DOCUMENT RESUME

ED 230 563 TM 830 192

AUTHOR	Wild, Cheryl L.; And Others
TITLE	Research Leading to the Revision of the Format of the Graduate Record Examinations Aptitude Test in October 1981.
INSTITUTION	Educational Testing Service, Princeton, N.J.; Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-82-55; GREB-80-1bP
PUB DATE	Nov 82
NOTE	99p.; The appendix and some tables may be marginally legible due to small print.
PUB TYPE	Reports - Research/Technical (143) -- Statistical Data (110)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Aptitude Tests; *College Entrance Examinations; Equated Scores; *Graduate Study; Higher Education; Research Methodology; Scoring; *Test Construction; *Test Format; Testing Problems; Testing Programs; Timed Tests
IDENTIFIERS	*Graduate Record Examinations; Test Content; Test Disclosure; *Test Revision

ABSTRACT

The research leading to the decisions to revise the Graduate Record Examination Aptitude Test (GRE) (beginning in October 1981) is reviewed. The issues discussed include the format of the test (the timing of each section and the number of sections, the content of the sections--especially the analytical section), the scoring procedure for the GRE, and test disclosure/administrative procedures. Pros and cons of the various revision alternatives are given. The format and content of the test will have direct implications on which equating methodologies are feasible, so the impact of the decisions on the different equating methodologies is considered throughout the discussion. A general summary of equating options is presented. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230563

GRE

GRADUATE RECORD EXAMINATIONS

RESEARCH LEADING TO THE REVISION OF THE FORMAT OF THE GRADUATE RECORD EXAMINATIONS APTITUDE TEST IN OCTOBER 1981

Cheryl L. Wild
Spencer S. Swinton
Madeline M. Wallmark

GRE Board Professional Report GREB No. 80-1bP
ETS Research Report 82-55

November 1982

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H.C. Weidenmiller
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 820 152

GRE BOARD RESEARCH REPORTS FOR GENERAL AUDIENCE

Altman, R. A. and Wallmark, M. M. A Summary of Data from the Graduate Programs and Admissions Manual. 74-1R, 1975.

Baird, L. L. An Inventory of Documented Accomplishments. 77-3R, 1979.

Baird, L. L. Cooperative Student Survey (The Graduates [$2.50 each], and Careers and Curricula). 70-4R, 1973.

Baird, L. L. The Relationship Between Ratings of Graduate Departments and Faculty Publication Rates. 77-2aR, 1980.

Baird, L. L. and Knapp, J. E. The Inventory of Documented Accomplishments for Graduate Admissions: Results of a Field Trial Study of Its Reliability, Short-Term Correlates, and Evaluation. 78-3R, 1981.

Burns, R. L. Graduate Admissions and Fellowship Selection Policies and Procedures (Part I and II). 69-5R, 1970.

Centra, J. A. How Universities Evaluate Faculty Performance: A Survey of Department Heads. 75-5bR, 1977. ($1.50 each)

Centra, J. A. Women, Men and the Doctorate. 71-10R, 1974. ($3.50 each)

Clark, M. J. The Assessment of Quality in Ph.D. Programs: A Preliminary Report on Judgments by Graduate Deans. 72-7aR, 1974.

Clark, M. J. Program Review Practices of University Departments. 75-5aR, 1977. ($1.00 each)

Clark, M. J and Centra, J. A. Conditions Influencing the Career Accomplishments of Ph.Ds. 76-2R, 1982.

DeVore, R. and McPeek, M. A Study of the Content of Three GRE Advanced Tests. 78-4R, 1982.

Donlon, T. F. Annotated Bibliography of Test Speededness. 76-9R, 1979.

Flaugher, R. L. The New Definitions of Test Fairness In Selection: Developments and Implications. 72-4R, 1974.

Fortna, R. O. Annotated Bibliography of the Graduate Record Examinations. 1979.

Frederiksen, N. and Ward, W. C. Measures for the Study of Creativity in Scientific Problem-Solving. 1978.

Hartnett, R. T. Sex Differences in the Environments of Graduate Students and Faculty. 77-2bR, 1981.

Hartnett, R. T. The Information Needs of Prospective Graduate Students. 77-8R, 1979.

Hartnett, R. T. and Willingham, W. W. The Criterion Problem: What Measure of Success in Graduate Education? 77-4R, 1979.

Knapp, J. and Hamilton, I. B. The Effect of Nonstandard Undergraduate Assessment and Reporting Practices on the Graduate School Admissions Process. 76-14R, 1978.

Lannholm, G. V. and Parry, M. E. Programs for Disadvantaged Students in Graduate Schools. 69-1R, 1970.

Miller, R. and Wild, C. L. Restructuring the Graduate Record Examinations Aptitude Test. GRE Board Technical Report, 1979.

Powers, D. E. and Swinton, S. S. Effects of Self-Study of Test Familiarization Materials for the Analytical Section of the GRE Aptitude Test. 79-9R, 1982.

Reilly, R. R. Critical Incidents of Graduate Student Performance. 70-5R, 1974.

Rock, D. and Werts, C. An Analysis of Time Related Score Increments and/or Decrements for GRE Repeaters across Ability and Sex Groups. 77-9R, 1979.

Rock, D. A. The Prediction of Doctorate Attainment in Psychology, Mathematics and Chemistry. 69-6aR, 1974.

Schrader, W. B. Admissions Test Scores as Predictors of Career Achievement in Psychology. 76-1aR, 1978.

Schrader, W. B. GRE Scores as Predictors of Career Achievement in [illegible]. 76-1bR, 1980.

Swinton, S. S. and Powers, D. E. A Study of the Effects of Special Preparation on GRE Analytical Scores and Item Types. 78-2R, 1982.

Wild, C. L. Summary of Research on Restructuring the Graduate Record Examinations Aptitude Test. 1979.

Wild, C. L. and Durso, R. Effect of Increased Test-Taking Time on Test Scores by Ethnic Group, Age, and Sex. 76-6R, 1979.

Wilson, K. M. A Study of the Validity of the Restructured GRE Aptitude Test for Predicting First-Year Performance in Graduate Study. 78-6R, 1982.

Wilson, K. M. The GRE Cooperative Validity Studies Project. 75-8R, 1979.

Wiltsey, R. G. Doctoral Use of Foreign Languages: A Survey. 70-14R, 1972. (Highlights $1.00, Part I $2.00, Part II $1.50).

Witkin, H. A.; Moore, C. A.; Oltman, P. K.; Goodenough, D. R.; Friedman, F.; and Owen, D. R. A Longitudinal Study of the Role of Cognitive Styles in Academic Evolution During the College Years. 76-10R, 1977 ($5.00 each).

Research Leading to the Revision of the Format of the Graduate Record Examinations Aptitude Test in October 1981

Cheryl L. Wild
Spencer S. Swinton
Madeline Wallmark

GRE Board Professional Report GREB No. 80-1bP

Copyright © 1982 by Educational Testing Service. All rights reserved.

TABLE OF CONTENTS

	Page
Introduction	1
Equating	1
Spiraling	1
Anchor Equating	2
Total Preequating	2
Section Preequating	2
Item Response Theory	2
Discussion of Format and Content Alternatives	3
Test Format	3
Test Disclosure/Administrative Procedures	3
Scoring Procedure	8
Content Considerations	9
Research Results	12
Timing	12
Practice Effects	22
Validity	27
Alternatives for the Analytical Measure	36
Summary of the Revisions	43
References	44
Attachment A. Supporting Conclusions	A-1

Appendix: Intercorrelations and reliabilities of six experimental sections (C19-C24) of analytical item types administered with GRE Aptitude Test Form K-ZGR3 in April 1980

LIST OF TABLES

		Page
Table 1.	Content of Operational and Variable Verbal Sections by Group	15
Table 2.	Scores on Verbal Questions in the Operational Versus Variable Section	15
Table 3.	Reliability and Speededness Information for 25- and 50-Minute Verbal Sections	16
Table 4.	Intercorrelations of 25- and 50-Minute Verbal Sections	17
Table 5.	Content of Operational and Variable Quantitative Sections by Group	19
Table 6.	Scores on Quantitative Questions in the Operational Versus Variable Section	19
Table 7.	Reliability and Speededness Information for 25- and 50-Minute Quantitative Sections	20
Table 8.	Intercorrelations of 25- and 50-Minute Quantitative Sections	21
Table 9.	Content of Operational and Variable Analytical Sections by Group	23
Table 10.	Scores on Analysis of Explanations in Operational Versus Variable Section	24
Table 11.	Reliability and Speededness for 25-Minute Analysis of Explanations Sections	24
Table 12.	Scores on Analytical Questions in Operational Versus Variable Sections	25
Table 13.	Reliability and Speededness for 25-Minute Logical Diagrams and Analytical Reasoning Sections	25
Table 14.	Intercorrelations of Analytical Variable Sections with the Total Analytical Score	26
Table 15.	Summary of Practice Effects for Verbal and Quantitative Item Types	28
Table 16.	Summary of Practice Effects for Analytical Item Types	29
Table 17.	Results of Pooled Departmental Analyses: First-Year Graduate GPA vs. GRE-V, GRE-Q, GRE-A, and Self-Reported (SR) Undergraduate GPA (UGPA)	31

LIST OF TABLES (continued)

Page

Table 18. Correlation of Scores on the Restructured GRE Aptitude Test, Separately and in Best-Weighted and Equally Weighted Composites, with First-Year Graduate GPA in Pooled Departmental Samples, Using Departmentally Standardized Variables: By Field..... 32

Table 19. Correlations of Verbal Item Types with UGPA by Location in the Test..... 34

Table 20. Correlation of Quantitative Item Types with UGPA by Location in the Test..... 35

Table 21. Correlation of Analytical Item Types with UGPA by Location in the Test..... 36

Table 22. Correlations of Analytical Items with UGPA..... 38

LIST OF FIGURES

		Page
Figure 1.	Content and Format of the Graduate Record Examinations Aptitude Test, October 1977 to September 1981.....	4
Figure 2.	Possible Formats for the GRE Aptitude Test.....	5
Figure 3.	Administrative Plans - Disclosure Options.....	7
Figure 4.	Versions of Test Administered in June 1980.....	13

Research Leading to the Revision of the Graduate Record Examinations Aptitude Test in October 1981

Introduction

Between September 1979 and September 1980, the Graduate Record Examinations (GRE) Board and its committees discussed a number of potential changes in the Aptitude Test as a result of recent research findings about the analytical measure and recent legislative action. Whenever changes are made in the content or format of a test, the possibility of scale discontinuity is introduced. It is generally considered better to introduce all changes at one time so that, if a discontinuity results, it will have a one-time impact, and all forms introduced after that time will be on the same scale. Changes were considered, therefore, that may be desirable from a more general perspective (for example, the possible change from formula to rights-only scoring) as well as changes that may be necessary due to legislation and research results.

This paper reviews the research leading to the decisions to revise the GRE Aptitude Test beginning in October 1981. Pros and cons of the various alternatives are also given. The issues discussed include the format of the test (the timing of each section and the number of sections), the content of the sections (especially the analytical section), the scoring procedure for the GRE Aptitude Test, and test disclosure/administrative procedures. The question of how to equate after October 1981 is beyond the scope of the paper. Several research projects (Kingston & Dorans, 1982a; Stewart, in progress) will have a direct impact on that issue. The format and content of the test, however, will have direct implications on which equating methodologies are feasible, so the impact of the decisions on the different equating methodologies is mentioned throughout the discussion and a general summary of the equating options is presented in the next section.

Equating

Changes in test content and/or format have direct implications for equating and were, therefore, made with equating requirements in mind. The following methods are the most likely ones for equating after October 1981.

(1) Spiraling -- This method is currently in use in the GRE Program. It involves simultaneously administering a new test and one or two previously equated tests to random groups of examinees. The assumption is made that any differences in statistics for the groups are attributable solely to the differences in test forms. So long as some large groups of examinees take the test without disclosure--as

is the case now on national test dates offered everywhere but New York--this method can continue.

(2) Anchor Equating -- This method is currently used in equating the Scholastic Aptitude Test. It was used for the GRE Aptitude Test until 1963 and again in October 1977 when the reformatted test was introduced. This method was also used in October 1981 to introduce the changes reported in this paper. It involves administering the same set of questions to examinees taking a new edition as had been administered to examinees taking a previous edition of the test. The performance of the group taking the new test and the performance of the group taking the earlier test is compared in terms of the common questions administered to both groups.

(3) Total Preequating -- This method involves administering a new test and an old test to the same examinees, but in counterbalanced order. Since the examinees are the same, any difference in statistics for the two tests can be attributed to form differences. This could be achieved for the GRE Aptitude Test by administering a new full-length verbal, quantitative, or analytical section along with the old test.

(4) Section Preequating -- This method involves administering each section of a future test to random groups of examinees taking a current test. By assuming that the groups are randomly equivalent and that certain stable relationships exist between the sections and the total test, a test can be equated before it is administered as a total test.

(5) Item Response Theory -- This method, currently used for the Test of English as a Foreign Language, involves equating through item statistics ("calibrated items") rather than section statistics as do the above methods. Items can be calibrated prior to administration together as a total test.

The current method of spiraling test forms to obtain random groups of examinees is not possible under full disclosure; that is, if all tests are disclosed, no old forms are available to administer with the new test. Spiraling could continue to be used in nondisclosure administrations, but, depending on the final decision about number of forms disclosed per year, it may not be sufficient to use this method alone. Application of any of the other four methods to the GRE Aptitude Test would be most appropriate if changes were to occur in the test format (see the next section for details). Since future legislation may require full disclosure, it may be advisable to change the GRE Aptitude Test format to allow for flexibility in the equating procedure even if no immediate change is made in equating.

Discussion of the Format and Content Alternatives

Test Format. The format of the GRE Aptitude Test between October 1977 and September 1981 is illustrated in Figure 1. The verbal and quantitative sections were each 50 minutes long while the two sections contributing to the analytical score and the variable section used for pretesting were 25 minutes long. After review of the sample test and several disclosed test booklets, examinees could deduce that the two 50-minute sections were always part of the operational test or "reported scores." Examinees could also observe that the fifth section was usually the section containing trial questions. For these reasons, examinees might be expected not to respond to section V in the same way as they respond to a section contributing to a reported score.* With the practice of using section V for estimating item difficulty for test construction purposes but not for equating, the possibility of responses to section V not being exactly like responses to an operational section presents some difficulties for test construction but no problem for equating by spiraling. However, if the variable section is to be used for equating, it is important for examinees to respond as though the variable section were an operational section of the test. One reason for restructuring the test to contain equally timed sections and varying the location of the variable section would be to assure that examinees responded to all sections in the same manner.

Besides the change to equally timed sections, the overall timing and number of sections was considered. Two variable sections would allow pretesting of more questions and would also allow use of the third method of equating (total preequating) discussed earlier. Thirty-minute rather than 25 minute sections may also be required for the anchor equating and item response theory equating options. Thirty-minute sections are easier administratively but lengthen the testing day considerably with the administration of the Advanced Tests in the afternoon. The possible formats considered for the restructured test are listed in Figure 2. Research results on timing and the impact on the final decision are reviewed later in this paper.

Test Disclosure/Administrative Procedures. During 1980-81 three test forms -- ones from the October, December, and February administrations -- were disclosed. April and June administrations were not offered in New York and were not disclosed. GRE Board policy has been that, when new services are offered to some examinees (for example, New York examinees), the new service will be offered to all examinees. Following this policy, it is not possible to equate tests by spiraling at disclosure administrations because either all forms would have to be disclosed or some examinees would not be able to order a copy of their test.

---

*A few test supervisors reported administrative problems with examinees wanting to leave after Section IV.

Figure 1

Content and Format of the
Graduate Record Examinations Aptitude Test
October 1977 to September 1981

Section	Item Types	Number of Questions	Time
I.	Verbal	80 (75)*	50 minutes
	a) Reading Comprehension	25 (22)	
	b) Antonyms	20 (22)	
	c) Analogies	18 (18)	
	d) Sentence Completions	17 (13)	
II.	Quantitative	55	50 minutes
	a) Quantitative Comparisons	30	
	b) Regular Mathematics	15	
	c) Data Interpretation	10	
III.	Analysis of Explanations	40 (36)**	25 minutes
IV.	Analytical	30	25 minutes
	a) Logical Diagrams	15	
	b) Analytical Reasoning -		
	Type I	10-11	
	Type II	4-5	
V.***	Trial Questions	Variable	25 minutes

*The number of verbal questions was decreased in October 1979 from 80 to 75 to decrease the speededness. Numbers on the left indicate the test content in October 1977, when the first restructured test was introduced.

**The number of questions in the analysis of explanations section was decreased from 40 to 36 in the new test edition introduced in June 1980.

***In some administrations, the trial section was section III, analysis of explanations section IV, and analytical reasoning section V.

Figure 2

Possible Formats for the GRE Aptitude Test

Number of Sections and Timing	Reported Score	Testing Time	Different From Original Time
(1) Same as original, 2 50-minute sections 3 25-minute sections	Verbal, Quantitative, and Analytical	175 min. or 2 hr. 55 min.	0
(2) 7 25-minute sections	Verbal, Quantitative, and Analytical	175 min. or 2 hr. 55 min.	0
(3) 3 25-minute sections 4 30-minute sections	Verbal, Quantitative, and Analytical	195 min. or 3 hr. 15 min.	20 min.
(4) 8 25-minute sections	Verbal, Quantitative, and Analytical	200 min. or 3 hr. 55 min.	25 min.
*(5) 7 30-minute sections	Verbal, Quantitative, and Analytical	210 min. or 3 hr. 30 min.	35 min.
(6) 4 25-minute sections 4 30-minute sections	Verbal, Quantitative, and Analytical	220 min. or 3 hr. 40 min.	45 min.
(7) 8 30-minute sections	Verbal, Quantitative, and Analytical	240 min. or 4 hrs.	65 min.
(8) 5 25-minute sections	Verbal and Quantitative	125 min. or 2 hr. 5 min.	-50 min.
(9) 5 30-minute sections	Verbal and Quantitative	150 min. or 2 hr. 30 min.	-25 min.
(10) 6 25-minute sections	Verbal and Quantitative	150 min. or 2 hr. 30 min.	-25 min.
(11) 6 30-minute sections	Verbal and Quantitative	180 min. or 3 hrs.	5 min.

*This timing was adopted for the October 1981 format of the GRE Aptitude Test.

The number of forms to be disclosed has a direct impact on the number of pretests to be developed, the type of equating to be used, and the number of administrations that can be offered in New York State. In determining the administrations to be disclosed in New York State for 1980-81, the GRE Program staff surveyed a number of New York institutions. The clear sentiment was that October, December, and February administrations in New York would be the most important from the institutional perspective. Because of the need to introduce multiple restructured forms in October 1981 (to have forms for handicapped, Monday, and special administrations), it was not possible to disclose a test administered in October 1981 if the Board was to be consistent with its policy of offering universal disclosure whenever a test is disclosed.

Depending on the number of forms developed, several long-term plans for disclosure were considered. Figure 3 presents three possible plans for GRE Aptitude Test disclosure. Plan A represents the least disclosure, with the goal of ultimately publishing two test forms a year while maintaining four secure forms. Plan B has a goal of ultimately publishing three test forms a year while maintaining four secure forms. Plan C represents a goal for full test disclosure, that is, disclosure of five Aptitude Test forms a year. These plans were based on the assumption that, given the test development schedule for 1980-81, the maximum number of tests that could be introduced in 1981-82 would be five. Each plan also assumes that during 1981-82 only two or three test editions would be disclosed (plus a sample test) while four or five new editions are introduced. In this way, some secure tests would be available for Monday, handicapped, and special administrations. If only two test editions a year are disclosed, the current equating procedure of spiraling could be used. If three test editions are disclosed a year, spiraling could continue to be used, but it would probably be desirable to supplement it periodically with a second method.

Depending on the number of examinees per year, it might be necessary to have two variable sections in the Aptitude Test for pretesting and equating if full disclosure were desired. Given the already long testing period, the small percent of examinees requesting copies of their actual test, and the reluctance of the Board to move to full disclosure without further information on its impact, this option was not considered feasible at this time.

Disclosure of two Aptitude Tests a year would mean offering the test in New York at only two of the five national administrations. Although this may be desirable during 1981-82 in order to build a pool of secure test forms, the long term consequences of such a cutback in services may not be desirable. For these reasons, the Board decided on the second option, with a goal of ultimately disclosing three test forms a year. This represents the same publication rate as planned for 1980-81.

It was recognized that the passage of laws could require full disclosure at all administrations nationally. The revised format and content of the Aptitude Test was intended to be compatible to such an event.

Figure 3

Administrative Plans

Disclosure Options

		A Least Disclosure	B	C Maximum Disclosure
1981-82	Number of New Tests Introduced	4	5	5
	Number of Tests Disclosed*	2	2	3
	Disclosure Dates	Dec./Feb. (Not Oct.)	Dec./Feb. (Not Oct.)	Dec./Feb./June (Not Oct.)
	Equating Methods Possible**	All	All	All/May Need 2 Procedures
	Secure at End of Year	2	3	2
	Number of Pretests for Next Year ***	60	60	75
1982-83	Number of New Tests Introduced	4	4	5
	Number of Tests Disclosed	2	3	5
	Disclosure Dates	Dec./Feb. (Not Oct.)	Any 3 Non-consecutive	All Sat.
	Equating Methods Possible	All	All/Many Need 2 Procedures	Only 3, 4, 5
	Secure at End of Year	4	4	2
	Number of Pretests for Next Year	30	45	75
1983-84	Number of New Tests Introduced	2	3	5
	Number of Tests Disclosed	2	3	5
	Disclosure Dates	Any 2 Admin.	Any 3 Admin.	All
	Equating Methods Possible	All	All but May Need 2 Procedures	Only 3, 4, 5
	Secure at End of Year	4	4	2
	Number of Pretests for Next Year	30	45	75

*Each plan assumes that only two or three test editions would be disclosed during 1981-82 in order to build up the inventory of secure tests for small administrations.

**These refer to the five methods listed earlier in this paper. This means that, based on disclosure alone, all options are possible. Discussions of format and content may preclude some options.

***This assumes that 15 pretests are needed to develop each new form (verbal, quantitative, and analytical). At least 500 people should take each pretest. If equating sections are used to introduce a new form, 45,000 people (9 sections x 5,000 examinees per section) would be needed. Eventually equating sections would become operational sections so no additional test development is assumed. If the number of examinees remains the same as in 1979-80 (unlikely with the change from six to five administrations), the full disclosure option could barely be done with one variable section.

Scoring Procedure. In April 1980, the Minority Graduate Education Committee, the Research Committee, and the GRE Board discussed the issues related to formula versus rights-only scoring. The GRE Aptitude Test prior to October 1981 was formula scored, that is, the score was based on the number of correct answers minus a fraction of the wrong answers. Formula scoring (coupled with formula directions) has the advantage of slightly reduced error variance in the reported scores while rights-only scoring (similarly coupled with rights directions) has the advantage that examinees know just what the best test-taking strategies are. There are also philosophical differences related to preferences for one or the other approach. The committees and the Board expressed a general preference for rights-only scoring, if the technical issues of the effect on reported scores could be solved satisfactorily.

The technical question to be addressed was, "Can scores based on rights-only scoring be made comparable to earlier test scores based on formula scoring?"

Several procedures that might be used in equating the GRE Aptitude Test if a change from formula to rights-only scoring was instituted were discussed:

1. Administer equating sections in section V of the April 1981 administration with directions for rights-only scoring. These same sections would be administered and used for anchor equating in October 1981. This procedure would require examinees to change from formula-scoring procedures to rights-only procedures within the same testing session.

2. Assume that a test administered with rights-only directions but formula scored would result in the same scores as the test administered with formula-score directions and formula scored. Anchor test equating could be done by administering an equating section in April 1981 with formula-score directions. The same equating section with rights-only directions would be given in October 1981. Formula scores on the October 1981 equating section (with the rights-only directions) would be used for equating.

3. Administer both the rights-only and formula-scored tests to the same people. Since the GRE Aptitude Test is so long, two full tests may be too much and cause fatigue. For that reason, an equating administration might be designed where each person takes either two verbal, two quantitative, or two analytical tests. Such a study would require a large number of examinees (about 6,000) who would probably have to be paid or given a free operational test.

Preliminary results from a study of rights-versus-formula scoring for the Scholastic Aptitude Test (Angoff & Schrader, 1981) suggested that examinees are able to change their behavior in accordance with new instructions (for example, option 1 above). Preliminary results also suggested that there is very little difference in the average formula scores of people who take the SAT with rights-only and formula-scoring

instructions (the assumption in option 2). The size of the possible jog in the scale as a result of making the assumption that formula scores are independent of the instructions was expected to be small.

Based on current evidence, it is impossible to guarantee that no discontinuity in the scale will occur if the scoring procedure is changed. However, it appears that any discontinuity could be minimized by introducing rights-only equating sections in the formula-scored tests. Thus, the decision about which scoring method is most appropriate for the GRE Aptitude Test appears to be a policy question rather than a technical one.

In making that decision, the Board also considered whether it was appropriate to offer the GRE Advanced Tests with formula-scoring instructions and, at the same time, offer the Aptitude Test with rights-only instructions. Although it may be feasible to equate formula-scored GRE Advanced Tests to rights-scored Advanced Tests using the assumption that formula scoring is invariant to the test instructions,* it was premature to suggest that rights-only scoring could be introduced for the Advanced Tests with a minimal impact on the scale. One might also argue that the basic differences between the GRE Aptitude Test and Advanced Tests make formula scoring more appropriate for the Advanced Tests. Questions on the Aptitude Test are not linked to any specific undergraduate curriculum and, therefore, all students might be expected to have some knowledge about all the questions. The Advanced Tests are constructed to cover a wide range of undergraduate curricula, and generally no single student would be expected to have covered all areas of the test. Thus, guessing might be considered less appropriate for the Advanced Tests. Given the differences between the GRE Aptitude Test and Advanced Tests and the preliminary finding that examinees can switch from rights to formula scoring, it seemed appropriate to introduce a change in the scoring method for the GRE Aptitude Test in October 1981 but not for the Advanced Tests. Thus, the final decision was to change the GRE Aptitude Test from formula to rights scoring.

Content Considerations. The potential changes discussed so far relate basically to the presentation of the test, but not to its substance. Since the introduction of the analytical measure in 1977, a number of research studies have been conducted to evaluate it. Research evidence has been mixed. Swinton and Powers (1982) found substantial short-term special preparation effects (estimated at 66 points) for the analytical measure. The logical diagrams and analysis of explanations item types appeared to be the primary contributors to this effect. Swinton, Wild, and Wallmark (1982) and Kingston and Dorans (1982b) also found a practice effect for analysis of explanations and logical diagrams. These types of questions are easier when they appear in section V than when they appear in section III or IV. Preliminary data also suggest that a larger proportion of repeaters gain more than 200 points on the analytical measure

---

* Angoff and Schrader also investigated the impact of this assumption for the College Board Achievement Test in Chemistry.

than on either the verbal or quantitative measure.* All of these results suggest that practice with some items in the analytical measure improves performance on the test. However, preliminary validity data, discussed at a later point in this report, appears positive.

The 1979-80 Guide to the Use of the Graduate Record Examinations stated that "the GRE Aptitude Test ... measures and yields separate scores for the basic verbal, quantitative, and analytical abilities a student has acquired over many years." Thus, the basic assumption about abilities measured by the Aptitude Test is that they cannot be influenced by short-term experience. These preliminary practice effect results raised an important policy question related to test content for the Board. Should the GRE Board continue to offer the analytical ability measure as originally structured in light of this evidence? Three alternative answers to this question were discussed:

1. Continue to offer the analytical measure (including the item types that have been show to be subject to practice effect), but provide practice materials and directions for special review to all examinees.

   The rationale for selecting this alternative was that the test was producing valid and useful results. Providing practice materials to all and encouraging the use of the materials by telling all examinees that a "practice effect" exists should equalize the effects of practice. The Research Committee was funding a study of preparation materials and the effects on analytical scores (Powers & Swinton, 1982). These materials could have been published before the 1981-82 testing year if the Board chose this alternative. However, the research study would not be able to tell whether coaching, in addition to studying the materials, would further increase scores. Neither could it tell whether providing study materials would eliminate the practice effect. Finally, it could not assess the value judgment about the "fairness" of testing in an area where short-term study before the test improves performance.

   Further, this option posed an equating problem. The spiraling method of equating (two old forms and one new form distributed alternately at the administration) is not affected by practice effect as long as the sections appear in the same location for everyone. Since the difficulty of some analytical item types is dependent on item location, some alternate equating methods (for

*After the October 1979 administration, of the repeaters since October 1977 with a 200-point gain in any score, 10 gained 200 or more points on verbal only, 31 gained 200 or more points on quantitative only, 69 gained 200 or more points on analytical only, 12 gained 200 or more points on two of the three scores, while 2 gained 200 or more points on all three.

example, item response theory methodology) would not be feasible for the analytical measure if it includes items subject to practice effects. In addition, anchor test equating would probably not be appropriate with the original test format due to the nonparallel nature of the two analytical sections.

2. Modify the test to reduce the practice effect.

   The rationale for selecting this alternative was that the test was producing valid and useful results, but that the existence of a significant practice effect in the absence of a method to insure that all examinees had the benefit of that practice could affect the fairness of the instrument. There are several ways of modifying the analytical measure that might result in a reduction of the practice effect:

   a. Use only the analytical reasoning item type and eliminate the other two (analysis of explanations and logical diagrams).

   b. Use the analytical reasoning item type and a new item type less affected by practice.

   c. Increase the time per item.

   Although these modifications may decrease the practice effect, they may change the meaning of the analytical score (in terms of validity as well as score scale). Thus, depending on the extent of the modification, analytical scores obtained after October 1, 1981, may not be comparable to those earned before the modification. The modification could have the advantage, however, of being amenable to a number of equating techniques, and the test might be designed to require only one variable section.

3. Drop the analytical measure as soon as possible.

   The rationale for selecting this alternative would be that the GRE Aptitude Test should only include items that are relatively free from practice or coaching effects. There are several advantages to this approach. First, the cost of producing new forms of the test would be reduced substantially. Second, the testing time at a single administration would be less. Finally, equating is much simpler for the verbal and quantitative sections than for the original analytical section. However, introducing the analytical measure had been an effort to broaden the skills measured by the test and the analytical score has shown preliminary evidence of predictive utility. Change that narrows the focus of the test may not be well received, especially in light of the validity evidence available to date.

The GRE Board discussed these alternative answers to the policy question and decided that modifying the test to reduce the practice effect was the preferred approach. However, before a final decision was made, additional research on the analytical measure was requested.

Research Results

To inform a decision concerning future format and content of the Aptitude Test, research on the following four topics was reviewed:

1. Appropriateness of 25-minute timings for sections with the then current test specifications (described in Figure 1);

2. Practice effect on verbal, quantitative, and analytical item types;

3. Validity of the restructured test and independence of the various scores; and

4. Alternate item types that might constitute the new analytical measure.

Timing. In June 1980, two test editions were administered. Each operational verbal and quantitative section was divided into two parallel halves. These 25-minute sections were then administered as the variable section in the alternate form. The 25-minute analytical sections were also administered in the variable section of the alternate version. This research design, illustrated in Figure 4, was also used to investigate practice effects for verbal, quantitative, and analytical item types and will be discussed from that perspective in a later section of this report.

In evaluating the timing of the 25-minute sections, three indicators of speededness were used: (1) the percent of examinees reaching the last question, (2) the percent of examinees completing three-fourths of the test, and (3) the number of items reached by 80 percent of the examinees. A test is generally considered unspeeded if virtually all the examinees complete three quarters of the test and if about 80 percent of the examinees reach the last item. Although a test meeting these standards may appear unspeeded from a technical perspective, it might still seem speeded to examinees. The GRE Aptitude Test is constructed to the extent possible with the most difficult questions toward the end of the separately timed sections. Thus, the statistical indications of speededness are confounded with item difficulty.

To evaluate the appropriateness of the 25-minute sections, the mean and standard deviations of scores on the operational and variable sections, the reliabilities of these sections, and the correlations of the operational and variable sections are presented. The mean scores on the operational sections provide an indication of the comparability of the study group to test takers for a full year. The mean and standard deviations for verbal, quantitative, and analytical scores for the 560,966 examinees tested between October 1, 1977, and June 30, 1979, are as follows:

Figure 4

Versions of Test Administered in June 1980

Test Edition ZGR1	Test Edition CGR1
I. Verbal - 50 minutes	I. Verbal - 50 minutes
II. Quantitative - 50 minutes	II. Quantitative - 50 minutes
III. Analysis of Explanations - 25 minutes	III. Analysis of Explanations - 25 minutes
IV. Logical Diagrams & Analytical Reasoning - 25 minutes	IV. Logical Diagrams & Analytical Reasoning - 25 minutes
V. 25 minutes of one of the following:	V. 25 minutes of one of the following:
a) 1st parallel half of Section I, CGR1	a) 1st parallel half of Section I, ZGR1
b) 2nd parallel half of Section I, CGR1	b) 2nd parallel half of Section I, ZGR1
c) 1st parallel half of Section II, CGR1	c) 1st parallel half of Section II, ZGR1
d) 2nd parallel half of Section II, CGR1	d) 2nd parallel half of Section II, ZGR1
e) Section III, CGR1	e) Section III, ZGR1
f) Section IV, CGR1	f) Section IV, ZGR1
	g) Section III, ZGR1 with 4 questions deleted

	Verbal	Quantitative	Analytical
Mean	480	518	495
Standard Deviation	129	135	130

Data in the following tables suggest that June test takers are somewhat below the full-year average. The tests will, therefore, appear more speeded for these groups. The means of the same questions in the operational versus the variable section will give some indication of the potential impact on the test score of changing the location of the subset of questions. More detail on practice effect is presented later.

Finally, the reliabilities of the variable sections and their correlations with the corresponding operational scores are presented to evaluate the appropriateness of using 25-minute sections for anchor equating. If the correlations and/or reliabilities are not sufficiently high, a longer section may be required.

Tables 1 through 4 present data for the four groups taking the 25-minute verbal variable sections. Table 1 presents summary statistics for each group on the operational verbal section and illustrates exactly how the content of the variable and operational sections varied by group. Table 2 contains mean scores for comparable items in the operational versus variable section. For example, the two groups (A and B) of 2,480 and 2,485 examinees who took the 39 questions of parallel form 1 in the operational section each had a mean of 15.42 while the 1,485 examinees in group C taking these questions in the variable section had a mean of 15.70.

The reliabilities and speededness information for each of the 25-minute sections and 50-minute operational sections are presented in Table 3 and correlations of 25-minute verbal sections with the operational verbal sections are reported in Table 4. The test appeared to be slightly speeded, which may be a problem for item response theory applications, especially when rights-only scoring is introduced (some item parameters cannot be estimated if a number of people respond randomly at the end of a test). The correlations of the variable sections with the operational sections are above .85, the suggested minimum for anchor test equating. It, therefore, appears that the timing for the new format of the verbal test should be increased, but no change in the number of questions in the test is suggested.

Tables 5 through 8 contain data for the four groups taking the 25-minute quantitative variable sections. Table 5 illustrates exactly how the content of the operational sections varied by group and presents summary statistics for each group on the operational quantitative section. Table 6 contains mean scores for the same items in the operational versus variable section.

Table 1

Content of Operational and Variable Verbal Sections by Group

Group	N	Operational Section			Variable Section
		Contents*	Mean	SD	Content
A	2,480	1 + 2	470	123	3
B	2,485	1 + 2	470	127	4
C	1,485	3 + 4	472	123	1
D	1,495	3 + 4	474	125	2

*Each operational section was divided into two parallel halves. See Figure 4.

Table 2

Scores on Verbal Questions in the Operational Versus Variable Section

		Operational (Embedded in 50 minutes)				25-Minute Variable Section			
Parallel Form	# Items	Group	Mean	SD	N	Group	Mean	SD	N
1	39	A	15.42	8.10	2,480	C	15.70	8.42	1,485
		B	15.42	8.14	2,485				
2	41	A	16.47	8.46	2,480	D	16.04	8.82	1,495
		B	16.37	8.81	2,485				
3	37	C	14.36	8.13	1,495	A	14.61	8.01	2,480
		D	14.24	8.05	1,485				
4	38	C	16.31	7.98	1,495	B	16.00	8.11	2,485
		D	16.24	7.92	1,485				

Table 3

Reliability and Speededness Information for
25- and 50-Minute Verbal Sections

Form	#Items	Rel.	SEM*	N	% Completing Test	% Completing 75%	#Items Reached by 80% of Examinees
Var. 1	39	.879	2.93	1,485	31.5	91	31
Var. 2	41	.880	3.05	1,495	30.6	91	36
Var. 3	37	.869	2.90	2,480	61.2	93	34
Var. 4	38	.877	2.85	2,485	45.9	96	34
Oper 1+2	80	.929	33	2,480	27.7	83	61
	80	.934	33	2,485	25.9	82	60
Oper 3+4	75	.929	33	1,485	40.2	90	65
	75	.931	33	1,495	42.8	90	67

*Variable section standard errors of measurement are presented in raw score units while operational section standard errors of measurement are presented in scaled score units.

Table 4

Intercorrelations of 25- and 50-Minute Verbal Sections

50-Minute Section	25-Minute Variable Section											
	1			2			3			4		
	r	(r)*	N	r	(r)	N	r	(r)	N	r	(r)	N
1 + 2							.881	(.981)	2,480	.890	(.983)	2,485
3 + 4	.872	(.965)	1,485	.881	(.973)	1,495						
	39 Items			41 Items			37 Items			38 Items		

*Correlations in parentheses are corrected for attenuation.

Table 5

Content of Operational and Variable Quantitative Sections by Group

Group	N	Operational Section			Variable Section
		Content*	Mean	SD	Content
A	2,495	1 + 2	498	132	3
B	2,480	1 + 2	496	131	4
C	1,485	3 + 4	498	130	1
D	1,495	3 + 4	492	132	2

*Each operational section was divided into two parallel halves. See Figure 4.

Table 6
Scores on Quantitative Questions in the Operational Versus Variable Section

Parallel Form	# Of Items	Operational				25-Minute Variable Section			
		Group	Mean	SD	N	Group	Mean	SD	N
1	27	A	11.97	5.62	2,495	C	11.57	5.55	1,485
		B	11.74	5.52	2,480				
2	28	A	12.66	5.54	2,495	D	12.11	5.46	1,495
		B	12.69	5.50	2,480				
3	27	C	12.66	6.17	1,485	A	11.84	6.38	2,495
		D	12.40	6.28	1,495				
4	28	C	12.43	5.99	1,485	B	12.65	5.92	2,480
		D	12.15	5.99	1,495				

Tables 7 and 8 report reliability, speededness, and intercorrelation information for each of the quantitative operational and variable sections. The information in Table 7 suggests that the 25-minute sections might be more speeded than one would like for item response theory equating, especially if rights-only scoring is introduced. The correlations of 50-minute operational and 25-minute variable sections are not as high as would be desirable for anchor equating. However, these correlations appear about as large as possible, given the current reliability. It, therefore, appears that the timing for the sections in the new format test should be 30 minutes and the number of questions increased to raise the reliability of the sections.

Since the analytical score was based on two separately timed sections, the operational section did not have to be divided into two variable sections as did the verbal and quantitative sections (see Figure 4). Table 9 illustrates the tests taken by the five groups with a variable analytical section. Groups A, B, and C took analysis of explanations (section III) questions in the variable section. Group C took a shortened (36-item) version of the 40-item operational analysis of explanations test. Groups D and E took logical diagrams and analytical reasoning questions (section IV) in the variable section. Section summary statistics, reliability, and speededness data for these three groups are presented in Tables 10 and 11. Section mean scores, reliability, and speededness data are presented in Tables 12 and 13. Intercorrelations of the operational and variable analytical sections are presented in Table 14.

The mean scores suggest that analysis of explanations questions are much easier after examinees have completed an earlier section of the same type of question. Decreasing the number of questions in the operational analysis of explanations section does not appear to have decreased the speededness in section III appreciably, even though the more able groups took the shorter test in section III. Note also that a variable analysis of explanations section appearing after an operational analysis of explanations section is less speeded regardless of the length.

In summary, the reliabilities of the 25-minute verbal sections and their correlations with the operational sections suggest that the tests could be used with anchor equating. However, the number of examinees completing the sections suggests that more time per item would be desirable. For the quantitative sections, the correlations of the operational and variable sections are not as high as desired for anchor equating. The standard error of measurement for the quantitative section is higher than for the verbal section. In order to raise the correlation and to increase reliability and therefore lower the standard error of measurement, it is suggested that five more questions be added to the operational quantitative section. Since the number of examinees completing the test suggests that it should not be more speeded, more time is also needed for the quantitative section. Thus, 30-minute sections are indicated for both the verbal and quantitative sections.

Table 7

Reliability and Speededness Information
for 25- and 50-Minute Quantitative Sections

	# Items	Rel	SEM*	N	% Completing Test	% Completing 75% Test	# Items Reached by 80% of Examinees
1	27	.784	2.58	1,485	66.3	98	24
2	28	.769	2.63	1,495	56.2	97	26
3	27	.840	2.55	2,495	50.1	97	24
4	28	.813	2.56	2,480	33.1	97	23
Oper 1&2	55	.877	46	2,495	50.1	97	50
		.872	47	2,480	49.4	98	50
Oper 3&4	55	.901	41	1,485	32.5	97	48
		.904	41	1,495	34.8	97	48

*Variable section standard errors of measurement are presented in raw score units while operational section standard errors of measurement are presented in scaled score units.

Table 8

Intercorrelations of 25- and 50-Minute Quantitative Sections

50-Minute Section	1			2			3			4		
	r	(r)*	N	r	(r)	N	r	(r)	N	r	(r)	N
1 & 2							.839	(.978)	2,495	.820	(.974)	2,480
3 & 4	.845	(1.01)	1,485	.807	(.965)	1,495						
		27 Items			28 Items			27 Items			28 Items	

*Correlations in parentheses are corrected for attenuation.

Table 9

Content of Operational and Variable
Analytical Sections by Group

Group	N	Operational Sections			Variable
		Content* Form/Section	Mean	SD	Content Form/Section
A	2,485	1 - III 1 - IV	491	129	2 - III
B	1,525	2 - III 2 - IV	497	131	1 - III
C	1,490	2 - III 2 - IV	493	132	1 - III (but only 36 of the 40 questions)
D	2,480	1 - III 1 - IV	489	133	2 - IV
E	1,475	2 - III 2 - IV	490	129	1 - IV

*See Figure 4 also.

Table 10

Scores on Analysis of Explanations in
Operational Versus Variable Section

		Operational Section				Variable Section			
Form	# Items	Group	Mean	SD	N	Group	Mean	SD	N
1	40	A	17.55	8.97	2,485	B	19.46	9.04	1,525
2	36	B	15.95	9.41	1,525	A	18.87	9.14	2,485
		C	15.92	9.50	1,490				
1′*	36		NA**	NA	NA	C	17.38	8.36	1,490

*1′ is the shortened version of form 1.

**NA=Data not available for this score.

Table 11

Reliability and Speededness for
25-Minute Analysis of Explanations Sections

Form	#Items/ Section	Rel	Sem	N	%Completing Test	%Completing 75%	#Items Reached by 80 % of Candidates
1	40-III	.861	3.35	2,485	70.2	93	35
	V	.864	3.34	1,525	83.9	97	40
2	36-III	.891	3.11	1,525	71.9	93	31
	III	.894	3.10	1,490	72.2	93	32
	V	.888	3.06	2,485	88.4	98	36
1′*	36-V	.857	3.17	1,490	86.6	98	36

*Form 1′ contains 36 of the 40 items in form 1.

Table 12

Scores on Analytical Questions in
Operational Versus Variable Sections

		Analytical Section				Variable Section			
Form	# Items	Group	Mean	SD	N	Group	Mean	SD	N
1-Total	30	D	15.69	7.40	2,480	E	16.03	7.19	1,475
2-Total	30	E	13.22	6.67	1,475	D	14.26	6.99	2,480
1-Logical Diagram	15	D	8.23	4.55	2,480	E	8.66	4.39	1,475
2-Logical Diagram	15	E	7.19	4.35	1,475	D	7.81	4.47	2,480
1-Analytical Reasoning	15	D	7.62	3.77	2,480	E	7.54	3.69	1,475
2-Analytical Reasoning	15	E	6.18	3.29	1,475	D	6.59	3.49	2,480

Table 13

Reliability and Speededness for
25-Minute Logical Diagrams and Analytical Reasoning Sections

Form	#Items	Section	Rel	Sem	N	%Completing Test	% Completing 75%	# Items Reached By 80% of Examinees
1	30	IV	.875	2.62	2,480	53.2	96	26
		V	.865	2.64	1,475	65.8	99	27
2	30	IV	.837	2.69	1,475	23.9	95	25
		V	.856	2.65	2,480	35.8	97	26
Total Analytical	70	III+IV	.926	36	2,480			
	66	III+IV	.923	36	1,475			

Table 14

Intercorrelations of Analytical Variable Sections with the Total Analytical Score

Total Analytical Score		Analysis of Explanations Variable Section									Analytical Reasoning and Logical Reasoning Variable Section					
		1			2			1'*			1			2		
Form	#Items	r	(r)**	N	r	(r)	N	r	(r)	N	r	(r)	N	r	(r)	N
1	70				.854	(.944)	2,485							.825	(.927)	2,480
2	66	.843	(.942)	1,525				.843	(.945)	1,490	.811	(.908)	1,475			

*Form 1' contains 36 of the 40 items in form 1.

**Correlations in parenthesis are corrected for attenuation.

Practice Effects. The impetus for studying practice effects came in October 1978 when it was first observed that the questions in the analytical section of the operational test were more difficult than anticipated from pretesting then in the fifth section of the test. The new form introduced in October 1978 was the first form developed from items pretested in the reformatted test, so this was the first chance to observe the change in difficulty. No similar differences in estimated item difficulty had been observed for verbal and quantitative questions.

The first investigation into the practice effect of analytical item types occurred using three test forms administered in April 1979. Preliminary results were reported to the Research Committee and the GRE Board in September 1979. The Research Committee raised a number of questions for further study: Do any of the other item types in the test show a practice effect? Should the content of the analytical measure be changed? Are there changes in test format that might reduce the practice effect? These questions resulted in a series of studies on practice effect for the GRE Aptitude Test (Swinton, Wild, & Wallmark, 1982).

Data in Tables 2, 6, 10, and 12 present some information on the mean scores of individuals on verbal, quantitative, and analytical operational and variable sections. The mean scores on the same questions in the operational versus variable section will give some indication of the potential impact of changing the location of the subset of questions. Note that practice effect is confounded with location effect; that is, we have no indication of how fatigue may be interacting with practice. Table 2 suggests that overall there is little practice effect on the verbal section--in two cases, items in section V appear easier and in two cases more difficult than the same items in the operational section. Table 6 shows little suggestion of practice effect on quantitative items, with three of the four means on section V below the means for the same questions in the operational section (II). However, Table 10 shows that mean scores on analysis of explanations items are about two and a half to three points higher in the variable section of the test. The mean score for logical diagrams and analytical reasoning items in the variable section is also higher than in the operational section by about one-half to one raw score point (Table 12).

A review of the results of practice effect studies for each item type is presented in Tables 15 and 16 using a more sophisticated technique than simple comparison of mean scores. Practice effect results were estimated using repeated measures analyses of covariance (see Swinton, Wild, & Wallmark (1982) for details). Note also in the following analyses that, due to the large numbers of examinees, many differences will be statistically significant but not practically significant.

These analyses suggest that the analysis of explanations item type exhibits the largest practice effect. The practice effect for logical diagrams is also large but slightly less. The supporting conclusions item type has a slightly larger practice effect than analytical reasoning item types, but this does not appear to be practically significant. Practice

Table 15

Summary of Practice Effects for Verbal and Quantitative Item Types

Item Type	n Items	n Individuals	Practice Effect	Practice Effect/Item	F
Reading Comprehension (explicit)	6	3970	- .06	-.01	6.44*
Reading Comprehension (explicit)	5	3976	- .25	-.05	96.35**
Reading Comprehension (inferential)	7	3970	.07	.01	4.68*
Reading Comprehension (inferential)	6	3976	- .37	-.06	122.39**
Sentence Completions	7	3970	.00	.00	0.00
Sentence Completions	8	3976	- .07	-.01	3.23
Analogies	8	3970	.18	.02	31.87**
Analogies	9	3976	.13	.01	16.61**
Antonyms	10	3970	.20	.02	28.81**
Antonyms	11	3976	.02	.00	43.99**
Regular Mathematics	8	3982	- .23	-.03	47.46**
Regular Mathematics	7	3976	- .01	-.00	0.04
Quantitative Comparisons	15	3982	- .26	-.02	33.60**
Quantitative Comparisons	15	3976	.26	-.02	30.25**
Data Interpretation	4	3982	- .13	-.03	24.98**
Data Interpretation	6	3976	.26	.04	61.11**

*$p < .05$

**$p < .01$

Table 16

Summary of Practice Effects for Analytical Item Types

Study	Item Type	n Items	n Individuals	Practice Effect	Practice Effect/Item	F
Study 1	Analysis of Explanations	40	2895	2.21	.06	372.87**
	Analysis of Explanations	40	2841	2.51	.06	570.46**
	Analysis of Explanations	40	2781	2.58	.06	552.04**
	Logical Diagrams	15	2767	.78	.05	213.42**
	Logical Diagrams	15	2814	.67	.04	113.18**
	Logical Diagrams	15	2853	.59	.04	139.04
	Analytical and Logical Reasoning	15	2867	.35	.02	34.60**
	Analytical and Logical Reasoning	15	2814	.29	.02	10.54**
	Analytical and Logical Reasoning	15	2853	.19	.01	23.13**
Study 2	Supporting Conclusions	25	829	.64	.03	19.00**
	Logical Reasoning	23	829	.12	.01	.97
Study 3	Logical Reasoning	4	3958	- .02	-.01	.73
	Analysis of Explanations	38	4005	2.44	.06	852.40**
	Logical Diagrams	15	3958	.53	.04	107.82**
	Analytical Reasoning	11	3958	.19	.02	21.25

** $p < .01$

effect for verbal and quantitative item types does not appear to be practically significant.

Validity. Any decision about the content and future of the analytical measure should include considerations of validity. Several aspects of validity might be considered -- the predictive validity or the relationship of the analytical test to success in graduate school, concurrent validity or the relationship of the test to other measures taken at the same time, and whether the analytical test is providing information that is independent of the current verbal and quantitative scores. In this section, validity research on the Aptitude Test in use between October 1977 and September 1981 will be reviewed. In addition, research on the reliability and validity of the future format of the analytical test will be presented.

Tables 17 and 18 contain data from Kenneth Wilson's study of the validity of the restructured GRE Aptitude Test (Wilson, 1982). Results for pooled analyses of English, education, history, sociology, mathematics, chemistry, computer science, and economics departments are presented. In comparing the validity coefficients of verbal, quantitative, and analytical scores and self-reported undergraduate grade-point average (UGPA) with graduate grade-point averages (GGPA) within each department, the correlation with the analytical score is the highest of the four correlations for four of the eight departments. The multiple correlation with GGPA improves a small amount when the analytical score is added to the other predictors in every case except mathematics. Thus, it appears that the analytical measure is valid and is adding to the information provided by the verbal and quantitative scores. However, in interpreting this data one should realize that validities for analytical scores may not be comparable to those for verbal and quantitative scores if direct selection is being based on verbal and quantitative scores but not on analytical scores.

In a review of the new analytical score after the first nine months (Miller & Wild, 1979), regression analyses suggested that the analytical score is providing information in multiple correlation with undergraduate grade-point averages that is different from that provided by the verbal and quantitative scores. Swinton and Powers (1980) also found a unique analytical factor defined mostly by analysis of explanations and logical diagrams item types. The analysis of explanations items were factorially complex, with responses B, C, and D loading on the analytical factor but with responses A and E more related to reading comprehension. Analytical reasoning type 2 (or logical reasoning) items were highly related to the reading comprehension factor. Only tentative conclusions could be made about the analytical reasoning type 1 items because they were at the end of a somewhat speeded section. However, there was some indication that these items were related to the analytical factor.

In the preliminary report of the confirmatory factor analysis, based on a group of social science majors (Rock, Werts & Grandy, 1982), the analysis of explanations item type was found to be the most factorially complex, including vocabulary, reading, and quantitative components. Logical diagrams had a larger unique variance component than analytical

Table 17

RESULTS OF POOLED DEPARTMENTAL ANALYSES: FIRST-YEAR GRADUATE GPA vs. GRE-V, GRE-Q, GRE-A, and SELF-REPORTED (SR) UNDERGRADUATE GPA (UGPA)

Variable	Verbal Fields					Quantitative Fields				
	English	Education	History	Sociology	ALL VERBAL	Chemistry	Mathematics	Computer Sci	Economics	ALL QUANT
Simple Correlation										
GRE-V	.230	.238	.365	.652	.281	.180	.209	.269	.071	.174
GRE-Q	.221	.214	.396	.405	.255	.265	.535	.215	.251	.286
GRE-A	.139	.332	.411	.268	.275	.300	.192	.428	.249	.298
SR-UGPA	.161	.188	.378	.394	.222	.288	.427	.219	.258	.286
Multiple Correlation										
V,Q	.273	.268	.446	.657	.318	.280	.535	.294	.251	.296
V,Q,A	.280	.337	.476	.699	.329	.324	.536	.428	.318	.340
V,Q,SR	.298	.296	.539	.723	.358	.390	.613	.335	.356	.384
V,Q,A,SR	.305	.355	.557	.778	.365	.415	.613	.452	.396	.415
Beta weights										
V	.173	.178	.231	.602	.211	.097	-.012	.218	-.003	.083
Q	.157	.136	.290	.097	.166	.230	.540	.128	.252	.256
V	.205	.050	.170	.752	.166	.006	.001	-.005	-.143	-.026
Q	.198	.043	.193	.216	.121	.147	.544	.017	.184	.188
A	-.088	.279	.214	-.333	.111	.218	-.022	.423	.260	.224
V	.183	.037	.149	.689	.145	.034	.910	.000	-.180	-.024
Q	.194	.029	.172	.309	.113	.125	.450	-.028	.195	.159
A	-.085	.268	.185	-.408	.102	.193	.022	.413	.233	.208
SR	.121	.117	.296	.356	.162	.262	.311	.151	.243	.240
Departmental Characteristics										
Mean $\bar{X}$ V	621	448	578	541	555	543	583	559	574	560
Q	510	450	522	537	500	662	733	709	653	679
A	563	463	556	567	535	598	614	609	614	607
SR	5.99	5.14	6.17	6.22	5.83	5.76	6.16	6.01	6.08	5.95
Grad GPA	3.66	3.66	3.41	3.68	3.61	3.45	3.44	3.42	3.38	3.42
Mean SD V	90	85	96	83	89	92	111	96	93	96
Q	99	88	100	101	97	88	80	81	89	86
A	101	100	88	108	99	84	92	91	83	87
SR	0.83	0.99	0.87	0.76	0.87	0.92	0.94	1.01	0.91	0.94
Grad GPA	0.34	0.36	0.59	0.42	0.41	0.38	0.42	0.58	0.48	0.45
No. depts.	16	11	8	6	41	20	7	10	13	50
N(with Apt/Grad GPA)	194	271	82	39	586	234	62	92	119	507
N(with Apt/Grad GPA and SR-UGPA)	177	251	80	38	546	200	60	91	106	457

Note 1: Regression results reflect relationships among variables standardized within departments prior to pooling. Values reported under "departmental characteristics" are means of departmental means and standard deviations, respectively, in the metric of the original variables, prior to standardization. Thus, for example, the mean GRE-Verbal mean for 16 English departments was 621, the mean GRE V standard deviation was 90.

Note 2: It is important to note that the coefficients for GRE V, Q, and A scores in Table 17 were derived from pooled departmental samples in which there were at least 5 students with a self-reported UGPA in addition to the Aptitude scores. Therefore, not all departments could be included in this analysis. The coefficients in Table 18 however are based on all departmental samples, i.e., to be included in the analyses, a department must have had a minimum of 5 students with scores on the restructured GRE Aptitude Test. Accordingly, due to sampling variations there are some discrepancies in the Aptitude coefficients reported in the two tables. These discrepancies however do not change the nature of the conclusions regarding the relative contributions of the respective measure.

Table 18

CORRELATION OF SCORES ON THE RESTRUCTURED GRE APTITUDE TEST, SEPARATELY AND IN BEST-WEIGHTED AND EQUALLY-WEIGHTED COMPOSITES, WITH FIRST-YEAR GRADUATE GPA IN POOLED DEPARTMENTAL SAMPLES, USING DEPARTMENTALLY-STANDARDIZED VARIABLES: BY FIELD

FIELD	No. of depts.	No. of cases	Simple correlation			Optimal weight *			V + Q + A	
			GRE-V	GRE-Q	GRE-A	GRE-V	GRE-Q	GRE-A	Optimal wts. R	Equal wts. r
VERBAL FIELDS										
ENGLISH	(18)	205	.208	.218	.136	.177	.199	-.080	.263	(.229)
EDUCATION	(12)	276	.226	.209	.322	.040	.042	.274	.326	(.304)
HISTORY	(10)	95	.352	.326	.362	.212	.128	.185	.428	(.425)
SOCIOLOGY	(7)	44	.635	.455	.326	.626	.312	-.228	.682	(.579)
[ALL VERBAL]	(47)	620	.269	.247	.267	.157	.118	.110	.317	(.317)
QUANTITATIVE FIELDS										
CHEMISTRY	(21)	239	.188	.273	.296	.015	.158	.203	.326	(.311)
MATHEMATICS	(7)	62	.209	.535	.192	.001	.544	-.022	.536	(.385)
COMPUTER SCIENCE	(11)	104	.245	.232	.425	-.028	.089	.408	.433	(.380)
ECONOMICS	(14)	124	.080	.206	.269	-.138	.134	.303	.313	(.237)
[ALL QUANTITATIVE]	(53)	529	.176	.280	.303	-.026	.184	.236	.344	(.316)

NOTE: These analyses were based on pooled, departmentally-standardized variables. For example, the coefficient of .208 for GRE-Verbal (V) shown for English departments is based on data for 205 students from 18 different departments; all variables were scaled as deviations in standard units from departmental means prior to pooling.

*Standard partial regression coefficients; note that negative coefficients occur for either GRE-V or GRE-A in several analyses, reflecting a "suppressor effect."

reasoning items, which appeared to add the least new information (they were more highly related to quantitative comparison items). Although the analytic factor in this study was highly correlated with the quantitative factor, it showed a higher correlation with self-reported undergraduate grades than the verbal or quantitative factors and had about the same correlation with self-reported undergraduate grades as the reading factor.

The impact of practice effect on test validity of the analytical measure cannot be directly measured. However, the correlations of the analytical scores in the operational and variable sections (that is, before and after practice) can be compared. Tables 19, 20, and 21 show the correlations of the same sets of items with self-reported undergraduate grade-point averages in both the operational and variable sections, thus giving some indication of validity before and after practice. The items were included in several operational groups, so some indication of group variability of the correlations is available. For the verbal, quantitative, and analytical measures, the correlations of the items in the variable section usually fall within the range of correlations exhibited by the same items in the operational section. This is true for the two item types exhibiting practice effect as well as for the other item types.

Thus, the evidence suggests that the analytical measure is positively related to graduate school grade-point averages and to self-reported undergraduate grade-point averages. The incremental validity of the analytical measure is difficult to assess, especially since the verbal and quantitative scores are probably being used more heavily than the analytical score in the admissions process. Finally, the impact of the alternative formats for the analytical measure on long term validity cannot be assessed.

Alternatives for the Analytical Measure. One of the options for the analytical measure is to change the content of the analytical sections. These might include the following:

1. Replace analysis of explanations items with a new item type called supporting conclusions.

2. Replace analysis of explanations with some combination of analytical reasoning type 1 and type 2 items.

3. Replace both analysis of explanations and logical diagrams items.

To evaluate the independence of the various analytical item types in relation to the verbal and quantitative measures, the concurrent validity of the various item types, and the reliability of various combinations of these item types, six special 25-minute subtests were administered in the April 1980 GRE Aptitude Test administration. These included the following:

1. 19 analytical reasoning type 1,

Table 19

Correlations of Verbal Item Types with UGPA by Location in the Test

Item Type	Form	# Items	Correlation Operational Section I Item Types by Group 1	2	3	4	5	6	7	Median	Range	Correlation Variable Section V	Difference Median*	Difference Range**
Sentence Completion	41	7	.20	.18	.18	.20	.19	.21		.195	.18-.21	.19	0	0
	42	10	.20	.25	.21	.23	.24	.26		.235	.20-.26	.19	-	-
	47	7	.27	.22	.23	.20	.22	.23	.25	.23	.20-.27	.21	-	0
	48	6	.21	.15	.17	.14	.15	.15	.19	.15	.14-.21	.19	+	0
Analogies	41	10	.18	.19	.18	.14	.19	.20		.19	.14-.20	.20	+	0
	42	10	.14	.19	.15	.16	.17	.19		.165	.14-.19	.14	-	0
	47	9	.22	.17	.19	.16	.16	.19	.17	.17	.16-.22	.16	-	0
	48	9	.10	.12	.14	.12	.16	.14	.15	.14	.12-.18	.15	+	0
Antonyms	41	10	.15	.19	.15	.14	.18	.19		.165	.14-.19	.20	+	+
	42	10	.20	.24	.20	.20	.24	.28		.22	.20-.28	.21	-	0
	47	10	.20	.18	.15	.16	.20	.18	.19	.18	.15-.20	.17	-	0
	48	12	.23	.19	.19	.18	.19	.22	.22	.19	.18-.23	.21	+	0
Reading Composition Explicit	41	8	.18	.20	.19	.18	.21	.25		.195	.18-.25	.23	+	0
	42	4	.17	.20	.20	.16	.21	.22		.20	.16-.22	.17	-	0
	47	4	.19	.11	.19	.15	.14	.19	.16	.16	.11-.19	.16	0	0
	48	5	.20	.15	.18	.19	.12	.19	.14	.18	.12-.20	.19	+	0
Reading Composition Inference	41	6	.14	.19	.20	.18	.20	.25		.195	.14-.25	.20	0	0
	42	7	.20	.24	.19	.20	.25	.21		.205	.19-.25	.18	-	-
	47	7	.23	.18	.20	.15	.17	.20	.19	.19	.15-.23	.16	-	0
	48	6	.19	.13	.18	.18	.19	.17	.17	.18	.13-.21	.21	-	0

*If the correlation for Section V equals the median correlation, a "0" is assigned. If it is less than the median, a "-" is assigned and if it is greater than the median, a "+" is assigned.

**If the correlation for Section V falls within the range of correlations, a "0" is assigned. If is below the range, a "-" is assigned and if it is above the range, a "+" is assigned.

Table 20

Correlation of Quantitative Item Types with UGPA by Location in the Test

Item Type	Form	# Items	Correlation Operational Section II by Group 1	2	3	4	5	6	7	Median	Range	Correlation Variable Section V	Difference Median*	Range**
Quantitative Comparison	43	15	.16	.17	.18	.15	.21	.18		.175	.15-.21	.15	-	0
	44	15	.16	.16	.17	.15	.17	.18		.165	.15-.18	.08	-	-
	49	15	.15	.17	.14	.10	.16	.16	.15	.15	.10-.17	.15	0	0
	50	15	.17	.16	.16	.11	.19	.19	.14	.16	.11-.19	.16	0	0
Regular Math	43	8	.16	.18	.16	.18	.16	.17		.175	.16-.18	.15	-	-
	44	7	.14	.13	.15	.13	.17	.14		.14	.13-.17	.05	-	-
	49	8	.15	.18	.17	.08	.17	.18	.14	.17	.08-.18	.18	+	0
	50	7	.11	.12	.11	.09	.09	.14	.11	.11	.09-.14	.16	+	+
Data Interpretation	43	4	.11	.14	.12	.12	.11	.14		.12	.11-.14	.10	-	-
	44	6	.11	.15	.12	.12	.13	.15		.125	.11-.15	.15	+	0
	49	4	.09	.10	.12	.10	.11	.10	.08	.10	.08-.12	.14	+	+
	50	6	.08	.10	.11	.10	.14	.14	.09	.10	.08-.14	.12	+	0

*If the correlation for Section V equals the median correlation, a "0" is assigned. If it is less than the median, a "-" is assigned and if it is greater than the median, a "+" is assigned.

**If the correlation for Section V falls within the range of correlations, a "0" is assigned. If is below the range, a "-" is assigned and if it is above the range, a "+" is assigned.

Table 21

Correlation of Analytical Item Types with UGPA by Location in the Test

Item Type	Form	# Items	Correlation Operational Section III or IV by Group 1	2	3	4	5	6	7	Median	Range	Correlation Variable Section V	Difference Median*	Range**
Analysis of Explanations	45	40	.22	.26	.24	.22	.25	.27		.245	.22-.27	.25	0	0
	51	36	.30	.24	.24	.19	.27	.22	.25	.24	.19-.30	.25	+	0
	53***	35	.22	.25	.23	.22	.25	.27		.24	.22-.27		-	0
	53	36										.22		
Logical Diagram	46	15	.21	.23	.22	.19	.22	.23		.22	.19-.23	.20	-	0
	52	15	.23	.18	.20	.15	.19	.22	.19	.19	.15-.23	.23	+	0
Analytical R-1	46	11	.21	.18	.22	.18	.21	.21		.21	.18-.22	.18	-	0
	52	11	.20	.16	.18	.13	.20	.18	.16	.18	.13-.20	.23	+	+
AR -2	46	4	.19	.21	.18	.19	.20	.21		.195	.19-.21	.18	-	0
-2	52	4	.19	.12	.18	.16	.18	.19	.15	.18	.12-.19	.23	+	+

*If the correlation for Section V equals the median correlation, a "0" is assigned. If it is less than the median, a "-" is assigned and if it is greater than the median, a "+" is assigned.

**If the correlation for Section V falls within the range of correlations, a "0" is assigned. If is below the range, a "-" is assigned and if it is above the range, a "+" is assigned.

***The scoring inadvertently included only 35 of the 36 questions for the operational section. This would have the effect of possibly lowering the reliability and the correlations.

2. 19 analytical reasoning type 2 (logical reasoning),

3. 5 analytical reasoning type 2 and 14 analytical reasoning type 1 (74 percent type 1 and 26 percent type 2),

4. 7 analytical reasoning type 2 and 12 analytical reasoning type 1 (63 percent type 1 and 37 percent type 2),

5. 25 supporting conclusions, and

6. 20 supporting conclusions.

Intercorrelations of these six variable sections with the operational sections for the total group are presented in Appendix Table 1 with simple correlations in the lower part of the correlation matrix and correlations corrected for attenuation in the upper part of the matrix. The analytical reasoning item type 1 appears less correlated with the verbal measure than either analytical reasoning type 2 or supporting conclusions. Supporting conclusions is more highly correlated with the analytical measure than with either the verbal or quantitative measures and is more highly related to the verbal measure than to the quantitative one. Logical reasoning is highly related to the verbal measure but not as highly related to the quantitative measure.

The second set of tables in the Appendix (Tables 2 through 7) presents the reliability of the six variable sections. The estimated reliability of doubling the first four variable sections and of adding the supporting conclusions variable sections to the original section IV, including the logical diagrams and analytical reasoning item types, are given below.

	#Items	Estimated Reliability
Analytical Reasoning Type 1	38	.88
Analytical Reasoning Type 2	38	.87
74% Type 1 + 26% Type 2	38	.89
63% Type 1 + 37% Type 2	38	.85
Supporting Conclusions (25 + Section IV)	55	.88
Supporting Conclusions (20 + Section IV)	50	.89

These estimated reliabilities for 25-minute sections suggest that two 30-minute analytical sections would be sufficiently reliable for a reported score for any of the above combinations.

Appendix Tables 8-13 present correlations of self-reported undergraduate grade-point averages with all the GRE Aptitude Test item types. Table 22 extracts the correlations of each of the six 25-minute analytical variable sections, the analysis of explanations section (III), the logical diagrams plus analytical reasoning section (IV), and the logical diagram items separately. Since the correlations are based on different samples, reliabilities for each set of items based on the total sample (correlations are based on a somewhat smaller sample because of the "no response" to the questions about grades) are presented in the far right column to help interpret the correlations. One of the original reasons for selecting analysis of explanations, logical diagrams, and analytical reasoning item types was a relatively high correlation with self-reported undergraduate grade-point averages. Table 22 suggests that differences in the correlations of the various item types with undergraduate grade-point average may not be any larger than might be expected within the same item type but based on separate samples. Correlations of undergraduate grade-point averages with analysis of explanation items and analytical reasoning type 2 items may be slightly higher than with the other three item types, when corrections are made for differences in reliability of the sections.

Correlations for the two supporting conclusions sections also have been obtained for the first time on a GRE population and suggest that supporting conclusions is about as valid as the other analytical item types. The two supporting conclusions sections, though, are not as reliable as the other item types. This may be a function of the item type, but may also be a function of the inexperience of test development staff in constructing the item. With pretesting of the items, one would expect a somewhat higher reliability.

The research concerning each item type and a description of the item type is presented below. Samples of the first four item types can be found in booklets containing full length tests given prior to October 1981 (Educational Testing Service, 1980; 1981)

1. Analysis of Explanations

These items are based on brief narratives establishing a situation in which an action is taken in order to have a specific effect. A later result, which may or may not be related to the action, is described in a brief statement. Each item is a statement that must be evaluated in terms of the situation and the result. The examinee must determine whether the statement is inconsistent with the situation or result, a possible adequate explanation of the result, deducible from the situation or result, relevant to or irrelevant to an explanation of the result.

Factor analytic studies of the GRE Aptitude Test have shown that this item type is factorially complex with options A and E highly related to reading comprehension. However, it does contribute to an analytical factor and correlates relatively highly with self-reported undergraduate grade-point averages. Directions are quite complex, which may partially explain the relatively large practice effect found for this item type. In

Table 22

Correlations of Analytical Items with UGPA

	#Questions	UGPA			
		4 Yr.	Major Field	Last 2 yrs.	Reliability Total Group *
Analytical Reasoning Type 1	19	.212	.213	.208	.782
Analysis of Explanations	40	.246	.250	.264	.881
Logical Diagrams + Analytical Reasoning (30 questions - Section IV)	30	.246	.266	.249	.876
Logical Diagrams	15	.229	.239	.231	.824
Analytical Reasoning Type 2	19	.256	.269	.251	.768
Analysis of Explanations	40	.267	.277	.268	.887
Logical Diagrams + Analytical Reasoning	30	.248	.267	.243	.878
Logical Diagrams	15	.216	.242	.221	.824
74% Analytical Reasoning-1, 26% Analytical Reasoning-2	19	.213	.204	.184	.797
Analysis of Explanations	40	.239	.270	.248	.879
Logical Diagrams + Analytical Reasoning	30	.223	.226	.235	.871
Logical Diagrams	15	.202	.205	.215	.820

*Reliabilities are estimated based on the total group taking each test (see appendix Tables 2-7). Examinees who did not respond to the questions on grades have been excluded in calculating the correlations.

Table 22 (continued)

Correlations of Analytical Items with UGPA

	# of Questions	UGPA			
		4 Yr.	Major Field	Last 2 Yrs.	Reliability Total Group*
63% Analytical Reasoning Type 1, 37% Type 2	19	.218	.234	.192	.742
Analysis of Explanations	40	.228	.235	.212	.881
Logical Diagrams + Analytical Reasoning	30	.245	.264	.223	.868
Logical Diagrams	15	.222	.238	.192	.816
Supporting Conclusions	25	.188	.194	.200	.679
Analysis of Explanations	40	.220	.232	.241	.885
Logical Diagrams + Analytical Reasoning	30	.239	.252	.260	.872
Logical Diagrams	15	.219	.216	.238	.824
Supporting Conclusions	20	.237	.228	.221	.672
Analysis of Explanations	30	.259	.272	.259	.885
Logical Diagrams + Analytical Reasoning	30	.254	.279	.266	.875
Logical Diagrams	15	.209	.237	.224	.822

*Reliabilities are estimated based on the total group taking each test (see Appendix Tables 2-7). Examinees who did not respond to the questions on grades have been excluded in calculating the correlations.

addition, this item type has been found to respond to coaching (Swinton & Powers, 1982).

2. Logical Diagrams

This item type presents several circle diagrams of possible relationships. Each item consists of three nouns and the student is asked to select the circle diagram that best characterizes the relationship of the three.

Factor analytic studies suggest that this item type is contributing to the analytical factor (Powers, Swinton, Thayer, & Yates, 1978; Swinton & Powers, 1980; Rock, Werts & Grandy, 1982). It does not correlate highly with verbal or quantitative scores. Correlations with self-reported undergraduate grade-point averages are slightly lower than for the analysis of explanations item type. This item type also has been found to have a relatively large practice effect (about two-thirds of that found in analysis of explanations) and to respond to coaching (Swinton & Powers, 1982).

3. Analytical Reasoning - Type 1

These items are based on a brief set of statements expressing relationships among abstract symbols (letters) or among sets of rules governing processes or procedures having few concrete referents. The examinee is asked to draw inferences from and sometimes critically assess those statements.

The correlation of these items with verbal and quantitative scores is relatively low. With undergraduate grade-point averages, the correlation is about the same as the logical diagrams and supporting conclusions item types. Analytical reasoning item types exhibit little practice effect, even though they appear at the end of a somewhat speeded section. The analytical reasoning item type is one of the two most highly rated by test development staff in terms of logical defensibility of the questions. One drawback of this item type is that it is time consuming for examinees.

4. Analytical Reasoning - Type 2

These items are based on brief arguments or statements presenting evidence or opinions. The questions require that the student recognize unstated presuppositions, logical flaws, methods of persuasion, conclusions logically following from given arguments, and the like.

Research on this item type for the Law School Admission Test program showed it to have a high criterion validity. It exhibits a high correlation with self-reported undergraduate grade-point average, but it is about as highly correlated with the verbal as with the analytical score. It has a relatively low correlation with the quantitative score. No practice effect has been found for this item type, and it is one of the two most highly rated analytical item types by the test development specialists

asked to rank the question types. However, this item type is difficult and time consuming to write.

5. Supporting Conclusions (See Attachment A).

A passage is presented with several numbered sentences. Following the passage are several statements. For each statement, the examinee must decide which of the numbered sentences are sufficient to supply all the information in each statement.

This item type correlates moderately with the verbal score and less with the quantitative score. It has a fairly high correlation with the analytical measure. A small practice effect was found for this item type. Test development staff have expressed concerns about the logical defensibility of this item type. Although it may have promise in the future, more study is needed before it is included as an operational item type.

Summary. The research on these five item types suggests that the two analytical reasoning item types are the most appropriate for a revised analytical section. The rankings for each item type based on selected criteria are presented below, where (+) indicates a favorable rating, (0) a neutral rating, and (-) an unfavorable rating. These ratings are relative only to the item types studied here. Note also that four of the five item types had already been screened (Miller & Wild, 1979) on some of the criteria listed below as well as some not included in this study (for example, item difficulty, efficiency, and face validity). For this reason, very few negative ratings have been given.

Item Type	Practice Effect	Independence		Correlation		Test Development Rating For Logical Defensibility
		V	Q	A	UGPA	
Analysis of Explanations	-	0	+	+	+	0
Logical Diagrams	-	0	+	0	0	0
*Analytical Reasoning - Type 1	+	+	+	0	0	+
*Analytical Reasoning - Type 2	+	-	+	+	+	+
Supporting Conclusions	0	0	+	+	0	0

*Item types proposed for the revised analytical measure.

The two analytical reasoning item types thus appear to be the most promising for a new analytical module. The two also compensate for each other -- analytical reasoning type 1 is more independent of the verbal measure while type 2 correlates more highly with self-reported undergraduate grade-point averages. In order not to inflate the correlation of the analytical and verbal scores, the new measure will include about 25 percent analytical reasoning type 2 and 75 percent analytical reasoning type 1 questions. These would be divided into two parallel separately timed sections (30 minutes each). Such a module could be expected to have a reliability of about .88 and intercorrelations with the verbal and quantitative measures in the range of .65 to .75.

Although it is expected that such a section will be feasible and useful, it represents a major change from the original analytical measure. In fact the item types in the revised measure constituted only 21 percent of the earlier measure. As a result, the predictive validity of the new analytical measure will have to be reassessed. Also due to the major change in the content, scores earned on the new test will not be comparable to those earned prior to the revision. This should not present a major problem to users, however, since the analytical score has been considered experimental to this time. Scores for the revised analytical measure will be rescaled in relation to the verbal and quantitative measures, rather than be linked to the old form.

Summary of the Revisions

The revisions to the GRE Aptitude Test discussed in this paper have to do with the timing and number of sections, the number of test forms to be disclosed, the scoring method, and content of the analytical measure. These revisions have been developed from the perspective of keeping as many equating options open as possible. The revisions implemented in October 1981, along with a brief summary of the rationale, are presented below:

1. The GRE Aptitude Test consists of seven 30-minute sections -- two verbal, two quantitative, two analytical, and one variable.

2. The analytical measure contains analytical reasoning item types, about three-fourths type 1 and one-fourth type 2.

3. Scoring has been changed from formula to rights scoring.

4. Five tests have been developed and two disclosed in 1981-82.

The change from varying length sections to equally timed sections was recommended to allow for flexibility in the equating procedure even with no immediate change in equating. Extending the time from 25 to 30 minutes was recommended to increase the time per item in the verbal sections and to increase the number of items and reliability of the quantitative sections.

Several options concerning the analytical section of the Aptitude Test were reviewed. One option, continuing to provide the analytical measure in its original format but providing practice materials to all examinees, was not considered feasible because there is no way to equalize the effects of practice and because short-term practice effects were large (50-60 points). A second option, eliminating the analytical measure, was not chosen because of the potential value of broadening the skills measured by the test and because the analytical measure has shown some evidence of validity. The final decision to include an analytical measure based on analytical reasoning item types 1 and 2 was based on an extensive review of five potential item types for inclusion (see the earlier section of this report).

The change from formula to rights-only scoring for the GRE Aptitude Test was based on the expressed preference of the GRE Board to move toward rights-only scoring and on preliminary evidence that the technical details can be worked out.

Finally, the decision about disclosure during 1981-82 directly influenced test development plans. Introduction of five test forms and publication of two (and a sample test) in 1981-82 allowed us to maintain some secure tests for small administrations and to move toward a goal of disclosing three tests a year (the disclosure rate for 1980-81).

References

Angoff, W. H., & Schrader, W. B. A study of alternative methods for equating rights scores to formula scores. ETS Research Report 81-8. Princeton, N.J.: Educational Testing Service, 1981.

Educational Testing Service. Graduate Record Examinations Aptitude Tests administered on January 12, April 26, and April 28, 1980. Princeton, N.J.: Author, 1980.

Educational Testing Service. Graduate Record Examinations Aptitude Test, Edition GR81-1. Princeton, N.J.: Author, 1981.

Educational Testing Service. Graduate Record Examinations Aptitude Test, Edition GR81-2. Princeton, N.J.: Author, 1981.

Educational Testing Service. Graduate Record Examinations Aptitude Test, Edition GR81-3. Princeton, N.J.: Author, 1981.

Kingston, N. M., & Dorans, N. J. The feasibility of using item response theory as a psychometric model for the GRE Aptitude Test. GRE Board Professional Report No. 79-12P. Princeton, N.J.: Educational Testing Service, 1982a.

Kingston, N. M., & Dorans, N. J. The effect of the position of an item within a test on item responding behavior; an analysis based on item response theory. GRE Board Professional Report No. 79-12bP. Princeton, N.J.: Educational Testing Service, 1982b.

Miller, R., & Wild, C. L. (Eds.) Restructuring the Graduate Record Examinations Aptitude Test. GRE Board Technical Report. Princeton, N.J.: Educational Testing Service, 1979.

Powers, D. E., & Swinton, S. S. Effects of self-study of test familiarization materials for the analytical section of the GRE Aptitude Test. GRE Board Research Report No. 79-9R. Princeton, N.J.: Educational Testing Service, 1982.

Powers, D. E., Swinton, S. S., Thayer, D., & Yates, A. A factor analytic investigation of seven experimental analytical item types. GRE Board Professional Report No. 77-1P. Princeton, N.J.: Educational Testing Service, 1978.

Rock D. A., Werts, C., & Grandy, J. Construct validity of the Graduate Record Examinations Aptitude Test across populations--an empirical confirmatory study. GRE Board Professional Report No. 78-1P. Princeton, N.J.: Educational Testing Service, 1982.

Stewart, E. E. Alternative equating methods for GRE. GRE Board Research Project 81-16, in progress.

Swinton, S. S., & Powers, D. E. A factor analytic study of the restructured GRE Aptitude Test. GRE Board Professional Report No. 77-6P. Princeton, N.J.: Educational Testing Service, 1980.

Swinton, S. S., & Powers, D. E. A study of the effects of special preparation on GRE analytical scores and item types. GRE Board Research Report No. 78-2R. Princeton, N.J.: Educational Testing Service, 1982.

Swinton, S. S., Wild, C. L., & Wallmark, M. Investigation of practice effects on types of questions in the Graduate Record Examinations Aptitude Test. GRE Board Professional Report No. 80-1cP. Princeton, N.J.: Educational Testing Service, 1982.

Wilson K. M. A study of the validity of the restructured GRE Aptitude Test for predicting first-year performance in graduate study. GRE Research Report No. 78-6R. Princeton, N.J.: Educational Testing Service, 1982.

ATTACHMENT A SUPPORTING CONCLUSIONS
ITEM TYPE

3 -15- 3 3 3 3 3 3 3 3 3

BEFORE TIME BEGINS PLEASE READ THE DIRECTIONS ON THIS PAGE AND THE NEXT.

SECTION III

Time—30 minutes

25 Questions

Directions: Each passage in this section consists of numbered sentences. Following each passage are several statements.

For each statement, you are to decide whether the information in the statement is explicitly supplied by the passage. If the information is NOT supplied, blacken space A on your answer sheet. If the information is supplied, decide whether B, C, or D gives the number(s) of the sentence(s) that make up the smallest group of sentences necessary to supply all of the information in the statement. If one of them does, it is the correct answer, and you should blacken the corresponding space on your answer sheet. If the information in the statement is explicitly supplied by the passage but none of the answer choices B, C, or D gives the smallest group of sentences necessary to supply the information in the statement, blacken space E on your answer sheet.

EXAMPLE

(1) Technology is the use of scientific knowledge to perform a practical task. (2) However, any such use requires us to divide the task into its parts. (3) Technology cannot be directly applied to large tasks like the construction of an airplane as a whole, but metallurgical technology, for example, can be used to develop alloys used in the parts of an airplane. (4) That is, the subdivisions of the task must correspond to established areas of scientific knowledge.

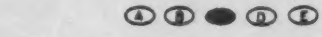
1. Statement: Metallurgical technology plays an important role in the design of airplanes.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1
(C) 1 and 4
(D) 3 and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

Explanation: The correct answer is A because the information in the statement, although it can be inferred, is not explicitly stated in the passage.

2. Statement: The use of scientific knowledge to perform a useful task requires that the task be divided into its components.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1
(C) 1 and 2
(D) 2, 3, and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.



Explanation: The correct answer is C because sentences 1 and 2 are the smallest group of sentences necessary to supply the information in the statement. Note that sentence 2 is not sufficient because the meaning of "such use" is given by sentence 1.

3. Statement: The use of technology requires that a task be divided into parts that correspond to categories of scientific knowledge.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1
(C) 2
(D) 1, 2, 3, and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

Explanation: The correct answer is E because sentences 1, 2, and 4 are the smallest group of sentences necessary to supply the information in the statement. Choices B and C contain too few sentences. Choice D contains too many sentences.

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

Set 1

(1) A recent ruling of the United States Supreme Court will have a significant effect on many municipal employees throughout the United States. (2) The ruling upheld the right of large municipalities to require that their municipal employees live within municipal boundaries. (3) Whether the decision will apply to small municipalities as well was left uncertain. (4) The United States Supreme Court's decision was based on the case of a Philadelphia fire fighter who was dismissed because he moved to New Jersey. (5) Local residency ordinances have been a controversial issue in many parts of the country, including New York, New Jersey, and Connecticut. (6) In New York City, a residence law requiring municipal employees to live within the city limits was repealed in 1962. (7) Legislation is pending to restore this requirement for new employees in New York City.

49. Statement: The United States Supreme Court ruling applies to large municipalities in the United States.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1
(C) 1 and 2
(D) 1, 2, and 3
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

50. Statement: The widely controversial issue of whether a large municipality can require its municipal employees to live within its boundaries has been ruled upon by the United States Supreme Court.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 2
(C) 2 and 5
(D) 1, 2, and 5
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

51. Statement: The United States Supreme Court's ruling supports proposed legislation that would introduce a change in residence regulations for new municipal employees in New York City.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 7
(C) 2 and 6
(D) 5, 6, and 7
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

52. Statement: The United States Supreme Court's decision may not have an effect on small municipalities.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 3
(C) 1, 2, and 3
(D) 1, 3, and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

53. Statement: As a result of a United States Supreme Court ruling, municipal employees of large municipalities must now live within the municipality in which they work.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 2
(C) 1, 2, and 3
(D) 1, 2, 3, and 5
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

GO ON TO THE NEXT PAGE.

Set 2

(1) Like Jupiter, Saturn has colored bands that mark zones of atmospheric motion. (2) Saturn's atmosphere, however, is far less turbulent than that of Jupiter. (3) So tranquil is Saturn that, during Pioneer's visit, scientists had difficulty locating an identifiable weather feature to track, frustrating plans of a precise clocking of the planet's rotation. (4) Saturn's most intriguing feature, however, is its rings. (5) Some scientists have thought them remnants of primordial material left over from the formation of the planet. (6) Other scientists thought that the rings were dense debris left from the breakup of a moon. (7) Recent measurements made of the deflection of Pioneer from its programmed path during its passage around Saturn, however, suggest that the mass of the rings is small. (8) The scientists monitoring the Pioneer flight regard the minimal deflection as an indication that the rings are predominantly composed of bits of ice, rather than of rock or metal.

54. Statement: The movement of a weather pattern can be used to establish the rate of rotation of a planet.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 3
(C) 1 and 2
(D) 1 and 3
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

55. Statement: Although Saturn and Jupiter have similar markings that result from atmospheric movement, the two planets are significantly different with respect to atmospheric turbulence.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 2
(C) 1 and 2
(D) 1, 2, and 3
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

56. Statement: Although the causes of weather patterns on Saturn are not understood, the rings of the planet are an even greater enigma.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 4
(C) 1, 2, and 4
(D) 1, 5, and 6
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

57. Statement: The planet Saturn has distinguishable areas of atmospheric motion.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1
(C) 2
(D) 1 and 2
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

58. Statement: Scientists' observations of the effect of Saturn's rings on the Pioneer spacecraft suggest a theory concerning the possible density and composition of the rings.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 5
(C) 5 and 6
(D) 6 and 7
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

GO ON TO THE NEXT PAGE.

Set 3

(1) A recent United States Supreme Court ruling held that citizens defamed by public officials cannot sue in federal courts for deprivation of their civil rights. (2) This Supreme Court decision did not affect the right to sue public officials for defamation in state courts. (3) However, such defamation suits are difficult to bring in many states, because of statutes or legal rules favoring public officials in civil suits. (4) The decision appeared to extend a recent trend toward denying persons complaining of constitutional violations by public officials access to federal courts. (5) The opinion of the Supreme Court was divided five to three in this decision.

59. Statement: Although the Supreme Court ruling did not limit access to state courts, some states do restrict the right to sue in their own courts.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 2
(C) 1 and 3
(D) 2 and 3
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

60. Statement: Before the recent Supreme Court decision, some citizens sued public officials in federal courts for violation of their civil rights because many states had laws favoring public officials in civil suits.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1 and 3
(C) 2 and 3
(D) 1, 2, and 3
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

61. Statement: The effects of the recent Supreme Court decision will be more strongly felt in some states than in others.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1, 2, and 3
(C) 1, 3, and 4
(D) 2, 3, and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

62. Statement: The Supreme Court decision apparently continues an existing trend which makes it more difficult for private citizens to defend their civil rights in suits against public officials.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 4
(C) 1 and 4
(D) 2, 3, and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

63. Statement: Five members of the Supreme Court supported a decision which seemed to continue a trend toward reducing citizens' access to court to contest violation of their civil rights by public officials.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 5
(C) 3 and 5
(D) 4 and 5
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

GO ON TO THE NEXT PAGE.

Set 4

(1) The Agriculture Department recently disclosed that two samples of pork were found to contain excess amounts of the toxic chemical PCB after the substance leaked into animal feed earlier this year. (2) PCB, or polychlorinated biphenyl, is a suspected cancer-causing agent. (3) The Assistant Secretary of Agriculture said the swine tissue samples were among more than 350 samples collected since September 17 to determine the extent of PCB contamination in swine, poultry, cattle, sheep, and egg products. (4) Exactly one per cent of the 200 pork samples tested were contaminated. (5) The samples were taken as part of the Department of Agriculture's recently accelerated monitoring program. (6) Federal regulations permit no more than three parts per million of PCB in food products. (7) Sampling of retail food products did not turn up any contamination.

64. Statement: The Department of Agriculture recently increased its efforts to detect polychlorinated biphenyl contamination in pork, poultry, cattle, sheep, and egg products.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1, 2, and 5
(C) 1, 2, 3, and 5
(D) 1, 2, 5, and 6
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

65. Statement: The Assistant Secretary of Agriculture reported that unacceptable levels of cancer-causing chemicals were found in one per cent of all pork tested since September 17.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1, 2, and 3
(C) 1, 2, 3, and 4
(D) 1, 2, 3, and 6
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

66. Statement: Samples of poultry, cattle, sheep, and egg products have been tested for PCB contamination since September 17.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 3
(C) 2 and 3
(D) 2, 3, and 4
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

67. Statement: The PCB contamination was discovered before it affected retail food products.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1, 2, and 7
(C) 1, 5, and 7
(D) 1, 2, 5, and 7
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

68. Statement: The Department of Agriculture recently announced that one per cent of the pork samples was found to contain more than 3 parts per million of PCB.

(A) The information in the statement is not explicitly supplied by the passage.
(B) 1, 4, and 6
(C) 2, 3, 4, and 6
(D) 3, 4, 5, and 6
(E) None of the above gives the smallest group of sentences necessary to supply the information in the statement.

GO ON TO THE NEXT PAGE.

GRE Aptitude Test Form K-ZGR3

Appendix TABLE 1

Intercorrelations of Verbal, Quantitative, Analytical and Pretest Scores, for Total Pretest Group

SAMPLE	Number of Items	VARIABLE	V	Q	A	C19	Raw Score Mean	Raw Score S.D.	Scaled Score Mean	Scaled Score S.D.
Pretest C19	80	Verbal		.533	.801	.636	32.99	14.62	468	124
N = 3,066	55	Quantitative	.485		.735	.697	24.96	11.15	494	129
Analytical Reasoning	70	Analytical	.740	.671		.781	35.90	15.80	482	132
	19	Pretest C19	.538	.583	.665		8.33	4.38		

SAMPLE	Number of Items	VARIABLE	V	Q	A	C20	Raw Score Mean	Raw Score S.D.	Scaled Score Mean	Scaled Score S.D.
Pretest C20	80	Verbal		.539	.808	.887	33.34	14.66	471	124
N = 3,029	55	Quantitative	.490		.735	.675	25.20	11.24	497	130
Logical Reasoning	70	Analytical	.748	.672		.903	36.74	16.02	489	133
	19	Pretest C20	.745	.560	.763		8.23	4.63		

SAMPLE	Number of Items	VARIABLE	V	Q	A	C21	Raw Score Mean	Raw Score S.D.	Scaled Score Mean	Scaled Score S.D.
Pretest C21	80	Verbal		.524	.801	.705	33.17	14.54	470	123
N = 3,042	55	Quantitative	.476		.748	.769	25.10	11.31	496	131
Analytical Reasoning and Logical Reasoning	70	Analytical	.739	.683		.871	36.40	15.52	487	129
	19	Pretest C21	.602	.650	.748		8.38	4.69		

SAMPLE	Number of Items	VARIABLE	V	Q	A	C22	Raw Score Mean	Raw Score S.D.	Scaled Score Mean	Scaled Score S.D.
Pretest C22	80	Verbal		.541	.807	.776	33.20	14.38	470	122
N = 3,046	55	Quantitative	.490		.741	.727	25.09	11.09	496	129
Analytical Reasoning and Logical Reasoning	70	Analytical	.744	.675		.864	36.31	15.60	486	130
	19	Pretest C22	.639	.592	.716		8.94	4.16		

SAMPLE	Number of Items	VARIABLE	V	Q	A	C23	Raw Score Mean	Raw Score S.D.	Scaled Score Mean	Scaled Score S.D.
Pretest C23	80	Verbal		.547	.804	.775	33.55	14.46	473	123
N = 3,044	55	Quantitative	.497		.745	.676	25.19	11.32	497	131
Supporting Conclusions	70	Analytical	.742	.681		.825	36.44	15.72	489	131
	25	Pretest C23	.611	.527	.654		9.06	4.59		

SAMPLE	Number of Items	VARIABLE	V	Q	A	C24	Raw Score Mean	Raw Score S.D.	Scaled Score Mean	Scaled Score S.D.
Pretest C24	80	Verbal		.524	.780	.792	33.27	14.51	471	123
N = 3,037	55	Quantitative	.476		.762	.686	24.93	11.28	494	131
Supporting Conclusions	70	Analytical	.721	.697		.860	36.33	15.85	486	132
	20	Pretest C24	.621	.532	.679		7.41	4.33		

NOTE: Values below the diagonal are observed correlations.
Values above the diagonal are correlations corrected for errors of measurement.

Appendix TABLE 2

Reliabilities and Standard Errors of Measurement for Reported Scores, Selected Operational Item Types, and Experimental and Augmented Analytical Reasoning

$K\text{-}ZGR3_{C19}$ — TOTAL GROUP — N = 3,245

VARIABLE	n	SECTION	SCORING FORMULA	RELIABILITY[a]	SE MEASUREMENT Raw	SE MEASUREMENT Scaled
Verbal Total	80	I	R-W/4	.919	4.18	35
Quantitative Total	55	II	R-.294W[c]	.896[b]	3.59	42
Quantitative Comparisons	30	II-a	R-W/3	.836	2.78	
Regular Mathematics + Data Interpretation	25	II-b	R-W/4	.794	2.28	
Analytical Total	70	III + IV	R-W/4	.929[b]	4.21	35
Analysis of Explanations	40	III	R-W/4	.881	3.34	
Logical Diagrams + Analytical Reasoning	30	IV	R-W/4	.876[b]	2.56	
Logical Diagrams	15	IV-a	R-W/4	.824	1.78	
Analytical Reasoning	15	IV-b	R-W/4	.772	1.84	
Experimental Analytical Reasoning	19	V	R-W/4	.782	2.05	
Augmented Analytical Reasoning	30	*	R-W/4	.853	2.59	

* Section IV-b, Analytical Reasoning (Type 1) plus Section V.

Appendix TABLE 3

Reliabilities and Standard Errors of Measurement for Reported Scores, Selected Operational Item Types, and Experimental and Augmented Logical Reasoning

$K\text{-}ZGR3_{C20}$ — TOTAL GROUP — N = 3,230

VARIABLE	n	SECTION	SCORING FORMULA	RELIABILITY[a]	SE MEASUREMENT Raw	SE MEASUREMENT Scaled
Verbal Total	80	I	R-W/4	.920	4.18	35
Quantitative Total	55	II	R-.294W[c]	.898[b]	3.59	42
Quantitative Comparisons	30	II-a	R-W/3	.837	2.78	
Regular Mathematics + Data Interpretation	25	II-b	R-W/4	.799	2.28	
Analytical Total	70	III + IV	R-W/4	.932[b]	4.18	35
Analysis of Explanations	40	III	R-W/4	.887	3.32	
Logical Diagrams + Analytical Reasoning	30	IV	R-W/4	.878[b]	2.54	
Logical Diagrams	15	IV-a	R-W/4	.824	1.77	
Analytical Reasoning	15	IV-b	R-W/4	.780	1.82	
Experimental Logical Reasoning	19	V	R-W/4	.768	2.24	
Augmented Logical Reasoning	23	*	R-W/4	.795	2.46	

* Section IV-b, Logical Reasoning (Type 2) plus Section V.

a) Adaptation of Kuder-Richardson formula (20).

b) See text.

c) $$K = \frac{\frac{30}{4} + \frac{25}{5}}{55 - (7.5 + 5)} = .294$$

Appendix TABLE 4

Reliabilities and Standard Errors of Measurement
for Reported Scores, Selected Operational Item Types,
and Experimental and Augmented Analytical and Logical Reasoning

K-ZGR3$_{C21}$ — TOTAL GROUP — N = 3,255

VARIABLE	n	SECTION	SCORING FORMULA	RELIABILITY[a]	SE MEASUREMENT Raw	SE MEASUREMENT Scaled
Verbal Total	80	I	R-W/4	.918	4.17	35
Quantitative Total	55	II	R-.294W[c]	.899[b]	3.59	42
Quantitative Comparisons	30	II-a	R-W/3	.841	2.77	
Regular Mathematics + Data Interpretation	25	II-b	R-W/4	.801	2.28	
Analytical Total	70	III + IV	R-W/4	.927[b]	4.20	35
Analysis of Explanations	40	III	R-W/4	.879	3.33	
Logical Diagrams + Analytical Reasoning	30	IV	R-W/4	.871[b]	2.55	
Logical Diagrams	15	IV-a	R-W/4	.820	1.78	
Analytical Reasoning	15	IV-b	R-W/4	.772	1.82	
Exp. Anal. Reas. + Logical Reas	19	V	R-W/4	.797	2.12	
Augmented Analytical Reasoning	25	*	R-W/4	.858	2.38	
Augmented Logical Reasoning	9	*	R-W/4	.613	1.44	

Appendix TABLE 5

Reliabilities and Standard Errors of Measurement
for Reported Scores, Selected Operational Item Types,
and Experimental and Augmented Analytical and Logical Reasoning

K-ZGR3$_{C22}$ — TOTAL GROUP — N = 3,235

VARIABLE	n	SECTION	SCORING FORMULA	RELIABILITY[a]	SE MEASUREMENT Raw	SE MEASUREMENT Scaled
Verbal Total	80	I	R-W/4	.917	4.18	35
Quantitative Total	55	II	R-.294W[c]	.895[b]	3.59	42
Quantitative Comparisons	30	II-a	R-W/3	.828	2.78	
Regular Mathematics + Data Interpretation	25	II-b	R-W/4	.799	2.27	
Analytical Total	70	III + IV	R-W/4	.927[b]	4.21	35
Analysis of Explanations	40	III	R-W/4	.881	3.34	
Logical Diagrams + Analytical Reasoning	30	IV	R-W/4	.868[b]	2.56	
Logical Diagrams	15	IV-a	R-W/4	.816	1.78	
Analytical Reasoning	15	IV-b	R-W/4	.764	1.84	
Exp. Anal. Reas. + Logical Rea	19	V	R-W/4	.742	2.12	
Augmented Analytical Reasoning	23	*	R-W/4	.816	2.25	
Augmented Logical Reasoning	11	*	R-W/4	.640	1.68	

* Section IV-b, Analytical Reasoning (Type 1) or Logical Reasoning (Type 2), plus same item types in Section V.

a) Adaptation of Kuder-Richardson formula (20).

b) See text.

c) $K = \dfrac{\frac{30}{4} + \frac{25}{5}}{55 - (7.5 + 5)} = .294$

Appendix TABLE 6

Reliabilities and Standard Errors of Measurement
for Reported Scores, Selected Operational Item Types,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C23}$ TOTAL GROUP N = 3,235

VARIABLE	n	SECTION	SCORING FORMULA	RELIABILITY[a]	SE MEASUREMENT Raw	Scaled
Verbal Total	80	I	R-W/4	.919	4.15	35
Quantitative Total	55	II	R-.294W[c]	.899[b]	3.59	42
Quantitative Comparisons	30	II-a	R-W/3	.840	2.77	
Regular Mathematics + Data Interpretation	25	II-b	R-W/4	.800	2.28	
Analytical Total	70	III + IV	R-W/4	.929[b]	4.19	35
Analysis of Explanations	40	III	R-W/4	.885	3.33	
Logical Diagrams + Analytical Reasoning	30	IV	R-W/4	.872[b]	2.55	
Logical Diagrams	15	IV-a	R-W/4	.824	1.77	
Analytical Reasoning	15	IV-b	R-W/4	.767	1.84	
Experimental Supporting Conclusions	25	V	R-W/4	.679	2.61	

Appendix TABLE 7

Reliabilities and Standard Errors of Measurement
for Reported Scores, Selected Operational Item Types,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C24}$ TOTAL GROUP N = 3,325

VARIABLE	n	SECTION	SCORING FORMULA	RELIABILITY[a]	SE MEASUREMENT Raw	Scaled
Verbal Total	80	I	R-W/4	.918	4.17	35
Quantitative Total	55	II	R-.294W[c]	.898[b]	3.59	42
Quantitative Comparisons	30	II-a	R-W/3	.840	2.77	
Regular Mathematics + Data Interpretation	25	II-b	R-W/4	.794	2.28	
Analytical Total	70	III + IV	R-W/4	.931[b]	4.18	35
Analysis of Explanations	40	III	R-W/4	.885	3.31	
Logical Diagrams + Analytical Reasoning	30	IV	R-W/4	.875[b]	2.55	
Logical Diagrams	15	IV-a	R-W/4	.822	1.78	
Analytical Reasoning	15	IV-b	R-W/4	.779	1.82	
Experimental Supporting Conclusions	20	V	R-W/4	.672	2.49	

a) Adaptation of Kuder-Richardson formula (20).

b) See text.

c) $K = \dfrac{\frac{30}{4} + \frac{25}{5}}{55 - (7.5 + 5)} = .294$

GRE Aptitude Test

Appendix TABLE 8a

Intercorrelations of Operational Subsections,
Self-Reported UGPAs, and
Experimental and Augmented Analytical Reasoning

$K\text{-}ZGR3_{C19}$ TOTAL GROUP N = 2,456

VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 Verbal Discrete											
2 Reading Comprehension	.688										
3 Quantitative Comparisons	.386	.442									
4 Regular Mathematics/Data Interpretation	.389	.451	.759								
5 Analysis of Explanations	.617	.666	.561	.554							
6 Logical Diagrams + Analytical Reasoning	.550	.598	.651	.619	.719						
7 UGPA - 4 Year	.229	.256	.186	.161	.246	.246					
8 UGPA - Major Field	.243	.260	.203	.187	.250	.266	.712				
9 UGPA - Last 2 Years	.239	.245	.201	.183	.264	.249	.679	.692			
10 EXPERIMENTAL Analytical Reasoning	.435	.501	.535	.559	.555	.626	.212	.213	.208		
11 AUGMENTED Analytical Reasoning	.489	.564	.616	.629	.644	.796	.223	.239	.226	.924	
Number of Items	55	25	30	25	40	30				19	23
Raw Score Mean	21.25	12.94	14.76	9.84	19.57	18.00				8.75	14.58
Raw Score S.D.	9.24	5.74	6.59	4.91	9.34	6.84				4.23	6.46

Appendix TABLE 8b

Intercorrelations of Verbal Item Types
Quantitative and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical Reasoning

K-ZGR3$_{C19}$ — TOTAL GROUP — N = 2,456

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Verbal-Sentence Completion													
2 Verbal-Analogies	.653												
3 Verbal-Antonyms	.653	.667											
4 RC-Explicit: Main Idea/Supporting Idea	.585	.516	.499										
5 RC-Intended Inference/Evaluation of Logic	.542	.475	.466	.606									
6 RC-Application, Style and Tone	.561	.514	.494	.597	.568								
7 Quantitative Total	.385	.401	.304	.447	.377	.383							
8 Analytical Total	.609	.559	.504	.608	.576	.569	.678						
9 UGPA - 4 Year	.212	.199	.194	.212	.220	.223	.187	.263					
10 UGPA - Major Field	.226	.208	.202	.225	.219	.223	.209	.276	.712				
11 UGPA - Last 2 Years	.228	.209	.192	.198	.227	.202	.207	.275	.679	.692			
12 EXPERIMENTAL Analytical Reasoning	.409	.411	.335	.462	.416	.402	.579	.630	.212	.213	.208		
13 AUGMENTED Analytical Reasoning	.466	.456	.374	.513	.468	.458	.661	.762	.223	.239	.226	.924	
Number of Items	17	18	20	9	8	8	55	70				19	30
Raw Score Mean	8.12	6.05	7.35	4.41	4.94	3.47	25.34	37.44				8.75	14.58
Raw Score S.D.	3.38	3.45	3.69	2.46	2.20	2.04	10.83	15.01				4.23	6.46

Appendix TABLE 8c

Intercorrelations of Quantitative Item Types,
Verbal and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical Reasoning

$K\text{-}ZGR3_{C19}$ TOTAL GROUP N = 2,456

	VARIABLE	1	2	3	4	5	6	7	8	9	10
1	Quantitative Comparisons										
2	Regular Mathematics	.747									
3	Data Interpretation	.559	.553								
4	Verbal Total	.442	.386	.415							
5	Analytical Total	.646	.583	.513	.708						
6	UGPA - 4 Year	.186	.159	.121	.260	.263					
7	UGPA - Major Field	.203	.177	.150	.270	.276	.712				
8	UGPA - Last 2 Years	.201	.167	.159	.261	.275	.679	.692			
9	EXPERIMENTAL Analytical Reasoning	.535	.526	.449	.500	.630	.212	.213	.208		
10	AUGMENTED Analytical Reasoning	.616	.595	.501	.561	.762	.223	.239	.226	.924	
	Number of Items	30	15	10	80	70				19	30
	Raw Score Mean	14.76	7.00	3.21	34.06	37.44				8.75	14.58
	Raw Score S.D.	6.59	3.41	2.13	13.81	15.01				4.23	6.46

Appendix TABLE 8d

Intercorrelations of Analytical Item Types
Verbal and Quantitative Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical Reasoning

K-ZGR3$_{C19}$ — TOTAL GROUP — N = 2,456

VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 Analysis of Explanations											
2 Analytical-Logical Diagrams	.631										
3 Analytical-Analytical Reasoning (Type 1)	.606	.551									
4 Analytical-Logical Reasoning (Type 2)	.477	.425	.417								
5 Verbal Total	.689	.534	.504	.466							
6 Quantitative Total	.594	.578	.610	.414	.472						
7 UGPA - 4 Year	.246	.229	.176	.191	.260	.187					
8 UGPA - Major Field	.250	.239	.210	.194	.270	.209	.712				
9 UGPA - Last 2 Years	.264	.231	.189	.170	.261	.207	.679	.692			
10 EXPERIMENTAL Analytical Reasoning	.555	.516	.587	.379	.500	.579	.212	.213	.208		
11 AUGMENTED Analytical Reasoning	.644	.593	.848	.440	.501	.661	.223	.239	.226	.924	
Number of Items	40	15	11	4	80	55				19	30
Raw Score Mean	19.57	10.15	5.85	1.87	34.06	25.34				8.75	14.58
Raw Score S.D.	9.34	3.99	3.00	1.21	13.81	10.83				4.23	6.46

Appendix TABLE 9a

Intercorrelations of Operational Subsections,
Self-Reported UGPAs, and
Experimental and Augmented Logical Reasoning

$K\text{-}ZGR3_{C20}$ TOTAL GROUP N = 2,454

	VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1	Verbal Discrete											
2	Reading Comprehension	.696										
3	Quantitative Comparisons	.411	.445									
4	Regular Mathematics/Data Interpretation	.402	.459	.758								
5	Analysis of Explanations	.628	.667	.552	.541							
6	Logical Diagrams + Analytical Reasoning	.583	.622	.662	.631	.731						
7	UGPA - 4 Year	.249	.251	.174	.142	.267	.248					
8	UGPA - Major Field	.256	.254	.185	.160	.277	.267	.722				
9	UGPA - Last 2 Years	.240	.226	.192	.155	.268	.243	.685	.685			
10	EXPERIMENTAL Logical Reasoning	.656	.682	.526	.538	.700	.674	.256	.269	.251		
11	AUGMENTED Logical Reasoning	.672	.699	.549	.554	.722	.724	.266	.277	.260	.975	
	Number of Items	55	25	30	25	40	30				19	23
	Raw Score Mean	21.43	13.06	14.79	10.07	20.08	18.22				8.59	10.46
	Raw Score S.D.	9.44	5.69	6.70	4.97	9.41	6.87				4.45	5.17

Appendix TABLE 9b

Intercorrelations of Verbal Item Types,
Quantitative and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Logical Reasoning

K-ZGR3$_{C20}$ — TOTAL GROUP — N = 2,454

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Verbal-Sentence Completion													
2 Verbal-Analogies	.641												
3 Verbal-Antonyms	.642	.671											
4 RC-Explicit: Main Idea/Supporting Idea	.598	.524	.486										
5 RC-Intended Inference/Evaluation of Logic	.570	.482	.477	.621									
6 RC-Application, Style and Tone	.570	.509	.490	.611	.560								
7 Quantitative Total	.435	.392	.317	.453	.369	.406							
8 Analytical Total	.638	.564	.520	.617	.579	.591	.675						
9 UGPA - 4 Year	.221	.201	.233	.222	.210	.215	.171	.277					
10 UGPA - Major Field	.232	.199	.241	.228	.217	.206	.185	.292	.722				
11 UGPA - Last 2 Years	.210	.190	.233	.204	.192	.182	.188	.275	.685	.685			
12 EXPERIMENTAL Logical Reasoning	.639	.567	.524	.598	.569	.584	.565	.740	.256	.269	.251		
13 AUGMENTED Logical Reasoning	.659	.578	.538	.616	.580	.601	.5[illegible]	.776	.266	.277	.260	.975	
Number of Items	17	18	20	9	8	8	55	70				19	23
Raw Score Mean	8.16	6.12	7.43	4.45	[illegible]	3.48	25.60	38.18				8.59	10.46
Raw Score S.D.	3.45	3.52	3.81	2.44	2.17	2.03	10.99	15.1[illegible]				4.45	5.17

Appendix TABLE 9c

Intercorrelations of Quantitative Item Types,
Verbal and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Logical Reasoning

$K\text{-}ZGR3_{C20}$ TOTAL GROUP N = 2,454

VARIABLE	1	2	3	4	5	6	7	8	9	10
1 Quantitative Comparisons										
2 Regular Mathematics	.749									
3 Data Interpretation	.573	.581								
4 Verbal Total	.457	.398	.419							
5 Analytical Total	.642	.577	.523	.723						
6 UGPA - 4 Years	.174	.141	.107	.270	.277					
7 UGPA - Major Field	.185	.156	.124	.275	.292	.722				
8 UGPA - Last 2 Years	.192	.148	.126	.253	.275	.685	.685			
9 EXPERIMENTAL Logical Reasoning	.526	.480	.481	.719	.740	.256	.269	.251		
10 AUGMENTED Logical Reasoning	.549	.494	.495	.738	.776	.266	.277	.260	.975	
Number of Items	30	15	10	80	70				19	23
Raw Score Mean	14.79	7.08	3.33	34.37	38.18				8.59	10.46
Raw Score S.D.	6.70	3.33	2.24	14.00	15.15				4.45	5.17

GRE Aptitude Test

Appendix TABLE 9d

Intercorrelations of Analytical Item Types,
Verbal and Quantitative Totals, Self-Reported UGPAs,
and Experimental and Augmented Logical Reasoning

$K\text{-}ZGR3_{C20}$ TOTAL GROUP N = 2,454

VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 Analysis of Explanations											
2 Analytical-Logical Diagrams	.657										
3 Analytical-Analytical Reasoning (Type 1)	.600	.578									
4 Analytical-Logical Reasoning (Type 2)	.503	.434	.412								
5 Verbal Total	.694	.573	.515	.507							
6 Quantitative Total	.583	.608	.609	.421	.486						
7 UGPA - 4 Year	.267	.216	.205	.188	.270	.171					
8 UGPA - Major Field	.277	.242	.212	.188	.275	.185	.722				
9 UGPA - Last 2 Years	.268	.221	.192	.174	.253	.188	.685	.685			
10 EXPERIMENTAL Logical Reasoning	.700	.589	.561	.500	.719	.565	.256	.269	.251		
11 AUGMENTED Logical reasoning	.722	.608	.581	.668	.738	.587	.266	.277	.260	.975	
Number of Items	40	15	11	4	80	55				19	23
Raw Score Mean	20.08	10.24	5.97	1.89	34.37	25.60				8.59	10.46
Raw Score S.D.	9.41	3.93	3.01	1.23	14.00	10.99				4.45	5.17

GRE Aptitude Test

Appendix TABLE 10a

Intercorrelations of Operational Subsections,
Self-Reported UGPAs, and
Experimental and Augmented Analytical and Logical Reasoning

$K\text{-}ZGR3_{C21}$ TOTAL GROUP N = 2,458

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
1 Verbal Discrete												
2 Reading Comprehension	.701											
3 Quantitative Comparisons	.397	.437										
4 Regular Mathematics/Data Interpretation	.395	.437	.767									
5 Analysis of Explanations	.623	.652	.564	.554								
6 Logical Diagrams + Analytical Reasoning	.560	.601	.660	.623	.706							
7 UGPA - 4 Year	.234	.225	.158	.126	.239	.223						
8 UGPA - Major Field	.245	.243	.171	.154	.270	.226	.706					
9 UGPA - Last 2 Years	.222	.228	.145	.132	.248	.235	.684	.699				
10 EXP. Analytical Reas. + Logical Reas.	.516	.567	.600	.596	.649	.700	.213	.204	.184			
11 AUGMENTED Analytical Reasoning	.477	.556	.633	.624	.646	.821	.199	.197	.189	.897		
12 AUGMENTED Logical Reasoning	.555	.540	.445	.456	.581	.620	.205	.200	.184	.678	.520	
Number of Items	55	25	30	25	40	30				19	25	9
Raw Score Mean	21.28	12.92	14.75	9.99	19.82	18.05				8.71	13.06	3.44
Raw Score S.D.	9.33	5.65	6.74	5.00	9.20	6.75				4.51	6.05	2.23

Appendix TABLE 10b

Intercorrelations of Verbal Item Types,
Quantitative and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical and Logical Reasoning

K-ZGR3$_{C21}$ TOTAL GROUP N = 2,458

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Verbal-Sentence Completion														
2 Verbal-Analogies	.658													
3 Verbal-Antonyms	.640	.670												
4 RC-Explicit: Main Idea/Supporting Idea	.590	.544	.503											
5 RC-Intended Inference/Evaluation of Logic	.568	.501	.460	.591										
6 RC-Application, Style and Tone	.573	.530	.468	.609	.578									
7 Quantitative Total	.400	.415	.300	.460	.326	.403								
8 Analytical Total	.613	.584	.507	.611	.555	.583	.684							
9 UGPA - 4 Year	.202	.196	.216	.212	.189	.170	.154	.251						
10 UGPA - Major Field	.222	.210	.212	.222	.204	.190	.174	.272	.706					
11 UGPA - Last 2 Years	.204	.187	.194	.200	.202	.178	.150	.261	.684	.699				
12 EXP. Analytical Reas. + Logical Reas.	.484	.474	.404	.529	.433	.491	.637	.725	.213	.204	.184			
13 AUGMENTED Analytical Reasoning	.456	.448	.357	.511	.426	.488	.670	.778	.199	.197	.189	.897		
14 AUGMENTED Logical Reasoning	.517	.493	.457	.487	.446	.453	.477	.645	.205	.200	.184	.678	.520	
Number of Items	17	18	20	9	8	8	55	70				19	25	9
Raw Score Mean	8.11	6.06	7.38	4.40	4.93	3.45	25.49	37.74				8.71	13.06	3.44
Raw Score S.D.	3.42	3.56	3.67	2.37	2.17	2.07	11.07	14.76				4.51	6.05	2.23

Appendix TABLE 10c

Intercorrelations of Quantitative Item Types,
Verbal and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical and Logical Reasoning

$K\text{-}ZGR3_{C21}$ TOTAL GROUP N = 2,458

VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 Quantitative Comparisons											
2 Regular Mathematics	.753										
3 Data Interpretation	.580	.572									
4 Verbal Total	.444	.382	.422								
5 Analytical Total	.653	.590	.522	.712							
6 UGPA - 4 Year	.158	.133	.079	.248	.251						
7 UGPA - Major Field	.171	.159	.102	.262	.272	.706					
8 UGPA - Last 2 Years	.145	.138	.086	.241	.261	.684	.699				
9 EXP. Analytical Reas. + Logical Reas.	.600	.577	.466	.578	.725	.213	.204	.184			
10 AUGMENTED Analytical Reasoning	.633	.607	.483	.547	.778	.199	197	.189	.897		
11 AUGMENTED Logical Reasoning	.445	.422	.384	.593	.645	.205	.200	.184	.678	.520	
Number of Items	30	15	10	80	70				19	25	9
Raw Score Mean	14.75	7.10	3.26	34.07	37.74				8.71	13.06	3.44
Raw Score S.D.	6.74	3.44	2.17	13.89	14.76				4.51	6.05	2.23

GRE Aptitude Test

Appendix TABLE 10d

Intercorrelations of Analytical Item Types,
Verbal and Quantitative Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical and Logical Reasoning

$K\text{-}ZGR3_{C21}$ TOTAL GROUP N = 2,458

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
1 Analysis of Explanations												
2 Analytical-Logical Diagrams	.629											
3 Analytical-Analytical Reasoning (Type 1)	.564	.529										
4 Analytical-Logical Reasoning (Type 2)	.500	.436	.417									
5 Verbal Total	.685	.548	.480	.504								
6 Quantitative Total	.595	.600	.586	.416	.472							
7 UGPA - 4 Year	.239	.202	.171	.169	.248	.154						
8 UGPA - Major Field	.270	.205	.168	.181	.262	.174	.706					
9 UGPA - Last 2 Years	.248	.215	.175	.175	.241	.150	.684	.699				
10 EXP. Analytical Reas. + Logical Reas.	.649	.577	.641	.456	.578	.637	.213	.204	.184			
11 AUGMENTED Analytical Reasoning	.646	.603	.877	.466	.547	.670	.199	.197	.189	.897		
12 AUGMENTED Logical Reasoning	.581	.469	.469	.763	.593	.477	.205	.200	.184	.678	.520	
Number of Items	40	15	11	4	80	55				19	25	9
Raw Score Mean	19.82	10.12	5.94	1.86	34.07	25.49				8.71	13.06	3.44
Raw Score S.D.	9.20	3.96	2.94	1.24	13.89	11.07				4.51	6.05	2.23

GRE Aptitude Test

Appendix TABLE 11a

Intercorrelations of Operational Subsections,
Self-Reported UGPAs, and
Experimental and Augmented Analytical and Logical Reasoning

K-ZGR3 022 — TOTAL GROUP — N = 2,488

	VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
1	Verbal Discrete												
2	Reading Comprehension	.677											
3	Quantitative Comparisons	.407	.439										
4	Regular Mathematics/Data Interpretation	.410	.441	.764									
5	Analysis of Explanations	.637	.655	.554	.561								
6	Logical Diagrams + Analytical Reasoning	.574	.606	.658	.640	.724							
7	UGPA - 4 Year	.241	.237	.161	.158	.228	.245						
8	UGPA - Major Field	.239	.224	.176	.159	.235	.264	.702					
9	UGPA - Last 2 Years	.195	.194	.161	.153	.212	.223	.681	.672				
10	EXP. Analytical Reas. + Logical Reas.	.554	.589	.540	.573	.637	.664	.218	.234	.192			
11	AUGMENTED Analytical Reasoning	.482	.543	.602	.624	.623	.799	.219	.236	.207	.842		
12	AUGMENTED Logical Reasoning	.601	.601	.470	.497	.641	.674	.206	.222	.178	.753	.548	
	Number of Items	55	25	30	25	40	30				19	23	11
	Raw Score Mean	21.39	12.97	14.83	10.04	19.75	17.93				9.24	12.26	4.64
	Raw Score S.D.	9.31	5.65	6.57	4.99	9.28	6.76				4.04	5.03	2.70

GRE Aptitude Test

Appendix TABLE 11b

Intercorrelations of Verbal Item Types,
Quantitative and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical and Logical Reasoning

$K\text{-}ZGR3_{C22}$ TOTAL GROUP N = 2,488

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Verbal-Sentence Completion														
2 Verbal-Analogies	.655													
3 Verbal-Antonyms	.648	.679												
4 RC-Explicit: Main Idea/Supporting Idea	.588	.516	.490											
5 RC-Intended Inference/Evaluation of Logic	.551	.466	.472	.611										
6 RC-Application, Style and Tone	.566	.478	.460	.600	.575									
7 Quantitative Total	.429	.399	.320	.445	.361	.386								
8 Analytical Total	.635	.569	.527	.602	.576	.569	.682							
9 UGPA - 4 Year	.218	.194	.222	.207	.204	.200	.169	.252						
10 UGPA - Major Field	.216	.201	.215	.199	.181	.199	.180	.266	.702					
11 UGPA - Last 2 Years	.177	.160	.174	.164	.176	.164	.167	.232	.681	.672				
12 EXP. Analytical Reas. + Logical Reas.	.542	.479	.439	.524	.493	.488	.590	.697	.218	.234	.192			
13 AUGMENTED Analytical Reasoning	.465	.429	.374	.493	.447	.445	.650	.750	.219	.236	.207	.842		
14 AUGMENTED Logical Reasoning	.587	.519	.481	.528	.502	.507	.512	.704	.206	.222	.178	.753	.548	
Number of Items	17	18	20	9	8	8	55	70				19	23	11
Raw Score Mean	8.22	6.06	7.38	4.43	4.96	3.45	25.62	37.53				9.24	12.26	4.64
Raw Score S.D.	3.43	3.45	3.70	2.39	2.17	2.04	10.90	14.92				4.04	5.03	2.70

Appendix TABLE 11c

Intercorrelations of Quantitative Item Types,
Verbal and Analytical Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical and Logical Reasoning

$K\text{-}ZGR3_{C22}$ TOTAL GROUP N = 2,488

VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 Quantitative Comparisons											
2 Regular Mathematics	.756										
3 Data Interpretation	.571	.579									
4 Verbal Total	.455	.393	.442								
5 Analytical Total	.642	.595	.536	.722							
6 UGPA - 4 Year	.161	.145	.131	.261	.252						
7 UGPA - Major Field	.176	.145	.137	.254	.266	.702					
8 UGPA - Last 2 Years	.161	.147	.117	.212	.232	.681	.672				
9 EXP. Analytical Reas. + Logical Reas.	.540	.532	.488	.616	.697	.218	.234	.192			
10 AUGMENTED Analytical Reasoning	.602	.591	.507	.548	.750	.219	.236	.207	.842		
11 AUGMENTED Logical Reasoning	.470	.447	.443	.652	.704	.206	.222	.178	.753	.548	
Number of Items	30	15	10	80	70				19	23	11
Raw Score Mean	14.83	7.10	3.31	34.22	37.53				9.24	12.26	4.64
Raw Score S.D.	6.57	3.43	2.16	13.75	14.92				4.04	5.03	2.70

Appendix TABLE 11d

Intercorrelations of Analytical Item Types,
Verbal and Quantitative Totals, Self-Reported UGPAs,
and Experimental and Augmented Analytical and Logical Reasoning

$K\text{-}ZGR3_{C22}$ TOTAL GROUP N = 2,488

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
1 Analysis of Explanations												
2 Analytical-Logical Diagrams	.653											
3 Analytical-Analytical Reasoning (Type 1)	.575	.539										
4 Analytical-Logical Reasoning (Type 2)	.487	.413	.390									
5 Verbal Total	.698	.568	.491	.477								
6 Quantitative Total	.592	.605	.606	.391	.486							
7 UGPA - 4 Year	.228	.222	.195	.163	.261	.169						
8 UGPA - Major Field	.235	.238	.203	.191	.254	.180	.702					
9 UGPA - Last 2 Years	.212	.192	.192	.139	.212	.167	.681	.672				
10 EXP. Analytical Reas. + Logical Reas.	.637	.565	.583	.428	.616	.590	.218	.234	.192			
11 AUGMENTED Analytical Reasoning	.623	.582	.874	.419	.548	.650	.219	.236	.207	.842		
12 AUGMENTED Logical Reasoning	.641	.537	.512	.738	.652	.512	.206	.222	.178	.753	.548	
Number of Items	40	15	11	4	80	55				19	23	11
Raw Score Mean	19.75	10.12	5.87	1.81	34.22	25.62				9.24	12.26	4.64
Raw Score S.D.	9.28	3.99	2.94	1.22	13.75	10.90				4.04	5.03	2.70

Appendix TABLE 12a

Intercorrelations of Operational Subsections,
Self-Reported UGPAs, and
Experimental Supporting Conclusions

$K\text{-}ZGR3_{C23}$ TOTAL GROUP N = 2,486

VARIABLE	1	2	3	4	5	6	7	8	9	10
1 Verbal Discrete										
2 Reading Comprehension	.686									
3 Quantitative Comparisons	.410	.472								
4 Regular Mathematics/Data Interpretation	.402	.466	.767							
5 Analysis of Explanations	.619	.662	.559	.550						
6 Logical Diagrams + Analytical Reasoning	.562	.620	.674	.634	.717					
7 UGPA - 4 Year	.224	.229	.167	.155	.220	.239				
8 UGPA - Major Field	.218	.233	.198	.184	.232	.252	.718			
9 UGPA - Last 2 Years	.211	.215	.193	.184	.241	.260	.675	.680		
10 EXPERIMENTAL Supporting Conclusions	.515	.592	.496	.509	.584	.591	.188	.194	.200	
Number of Items	55	25	30	25	70	30				25
Raw Score Mean	21.36	13.14	14.79	10.03	19.94	18.01				9.42
Raw Score S.D.	9.31	5.73	6.78	4.98	9.42	6.86				4.47

GRE Aptitude Test

Appendix TABLE 12b

Intercorrelations of Verbal Item Types,
Quantitative and Analytical Totals, Self-Reported UGPAs,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C23}$ TOTAL GROUP N = 2,486

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
1 Verbal-Sentence Completion												
2 Verbal-Analogies	.668											
3 Verbal-Antonyms	.646	.676										
4 RC-Explicit: Main Idea/Supporting Idea	.597	.526	.489									
5 RC-Intended Inference/Evaluation of Logic	.561	.476	.446	.611								
6 RC-Application, Style and Tone	.566	.515	.482	.619	.569							
7 Quantitative Total	.422	.402	.319	.469	.384	.429						
8 Analytical Total	.617	.569	.507	.623	.574	.586	.684					
9 UGPA - 4 Year	.217	.166	.211	.214	.191	.186	.171	.246				
10 UGPA - Major Field	.213	.159	.207	.216	.193	.193	.203	.260	.718			
11 UGPA - Last 2 Years	.201	.157	.201	.194	.185	.175	.201	.269	.675	.680		
12 EXPERIMENTAL Supporting Conclusions	.507	.458	.398	.541	.471	.509	.533	.632	.188	.194	.200	
Number of Items	17	18	20	9	8	8	55	70				25
Raw Score Mean	8.22	6.15	7.29	4.50	5.00	3.52	25.56	37.82				9.42
Raw Score S.D.	3.45	3.52	3.60	2.44	2.18	2.06	11.11	15.12				4.47

TE Aptitude Test

Appendix TABLE 12c

Intercorrelations of Quantitative Item Types
Verbal and Analytical Totals, Self-Reported UGPAs,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C23}$ TOTAL GROUP N = 2,486

VARIABLE	1	2	3	4	5	6	7	8	9
1 Quantitative Comparisons									
2 Regular Mathematics	.757								
3 Data Interpretation	.578	.574							
4 Verbal Total	.469	.417	.405						
5 Analytical Total	.654	.591	. 23	.716					
6 UGPA - 4 Year	.167	.142	.129	.245	.246				
7 UGPA - Major Field	.198	.175	.145	.242	.260	.718			
8 UGPA - Last 2 Years	.193	.179	.138	.230	.269	.675	.680		
9 EXPERIMENTAL Supporting Conclusions	.496	.459	.449	.590	.632	.188	.194	.200	
Number of Items	30	15	10	80	70				25
Raw Score Mean	14.79	7.10	3.30	34.40	37.82				9.42
Raw Score S.D.	6.78	3.40	2.19	13.86	15.12				4.47

Appendix TABLE 12d

Intercorrelations of Analytical Item Types
Verbal and Quantitative Totals, Self-Reported UGPAs,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C23}$ TOTAL GROUP N = 2,486

VARIABLE	1	2	3	4	5	6	7	8	9	10
1 Analysis of Explanations										
2 Analytical-Logical Diagrams	.654									
3 Analytical-Analytical Reasoning (Type 1)	.566	.545								
4 Analytical-Logical Reasoning (Type 2)	.501	.451	.420							
5 Verbal Total	.688	.575	.485	.479						
6 Quantitative Total	.590	.611	.598	.448	.496					
7 UGPA - 4 Year	.220	.219	.188	.164	.245	.171				
8 UGPA - Major Field	.232	.216	.217	.181	.242	.203	.718			
9 UGPA - Last 2 Years	.241	.238	.206	.175	.230	.201	.675	.680		
10 EXPERIMENTAL Supporting Conclusions	.584	.498	.512	.433	.590	.533	.188	.194	.200	
Number of Items	40	15	11	4	80	55				25
Raw Score Mean	19.94	10.15	5.87	1.85	34.40	25.56				9.42
Raw Score S.D.	9.42	4.02	2.96	1.24	13.86	11.11				4.47

Appendix TABLE 13a

Intercorrelations of Operational Subsections,
Self-Reported UGPAs, and
Experimental Supporting Conclusions

$K\text{-}ZGR3_{C24}$ TOTAL GROUP N = 2,443

VARIABLE	1	2	3	4	5	6	7	8	9	10
1 Verbal Discrete										
2 Reading Comprehension	.673									
3 Quantitative Comparison	.379	.447								
4 Regular Mathematics/Data Interpretation	.393	.459	.781							
5 Analysis of Explanations	.608	.654	.578	.584						
6 Logical Diagrams + Analytical Reasoning	.520	.589	.677	.660	.731					
7 UGPA - 4 Year	.275	.287	.185	.184	.259	.254				
8 UGPA - Major Field	.256	.284	.191	.207	.272	.279	.726			
9 UGPA - Last 2 Years	.251	.258	.163	.182	.259	.266	.684	.691		
10 EXPERIMENTAL Supporting Conclusions	.524	.586	.513	.517	.628	.607	.237	.228	.221	
Number of Items	55	25	30	25	70	30				20
Raw Score Mean	21.35	12.92	14.75	9.97	19.94	17.89				7.68
Raw Score S.D.	9.32	5.77	6.76	5.00	9.36	6.94				4.25

Appendix TABLE 13b

Intercorrelations of Verbal Item Types,
Quantitative and Analytical Totals, Self-Reported UGPAs,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C2/}$ TOTAL GROUP N = 2,443

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
1 Verbal-Sentence Completion												
2 Verbal-Analogies	.637											
3 Verbal-Antonyms	.640	.668										
4 RC-Explicit: Main Idea/Supporting Idea	.593	.521	.493									
5 RC-Intended Inference/Evaluation of Logic	.567	.450	.443	.633								
6 RC-Application, Style and Tone	.563	.489	.461	.623	.587							
7 Quantitative Total	.404	.384	.284	.446	.388	.397						
8 Analytical Total	.598	.540	.472	.597	.566	.575	.702					
9 UGPA - 4 Year	.240	.229	.257	.253	.231	.260	.195	.276				
10 UGPA - Major Field	.238	.203	.231	.246	.241	.247	.209	.295	.726			
11 UGPA - Last 2 Years	.226	.207	.226	.218	.228	.222	.181	.281	.684	.691		
12 EXPERIMENTAL Supporting Conclusions	.533	.449	.399	.526	.494	.495	.544	.664	.237	.228	.221	
Number of Items	17	18	20	9	8	8	55	70				20
Raw Score Mean	8.14	6.17	7.33	4.48	4.95	3.37	25.46	37.70				7.68
Raw Score S.D.	3.49	3.45	3.69	2.47	2.17	2.04	11.14	15.18				4.25

Appendix TABLE 13c

Intercorrelations of Quantitative Item Types
Verbal and Analytical Totals, Self-Reported UGPAs,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C24}$ TOTAL GROUP N = 2,443

VARIABLE	1	2	3	4	5	6	7	8	9
1 Quantitative Comparisons									
2 Regular Mathematics	.766								
3 Data Interpretation	.591	.574							
4 Verbal Total	.440	.406	.406						
5 Analytical Total	.666	.623	.540	.691					
6 UGPA - 4 Year	.185	.169	.154	.305	.276				
7 UGPA - Major Field	.191	.188	.179	.290	.295	.726			
8 UGPA - Last 2 Years	.163	.172	.148	.276	.281	.684	.691		
9 EXPERIMENTAL Supporting Conclusions	.513	.474	.444	.596	.664	.337	.228	.221	
Number of Items	30	15	10	80	70				20
Raw Score Mean	14.75	7.07	3.27	34.14	37.70				7.68
Raw Score S.D.	6.76	3.44	2.17	13.86	15.18				4.25

Appendix TABLE 13d

Intercorrelations of Analytical Item Types,
Verbal and Quantitative Totals, Self-Reported UGPAs,
and Experimental Supporting Conclusions

$K\text{-}ZGR3_{C24}$ TOTAL GROUP N = 2,443

VARIABLE	1	2	3	4	5	6	7	8	9	10
1 Analysis of Explanations										
2 Analytical-Logical Diagrams	.655									
3 Analytical-Analytical Reasoning (Type 1)	.588	.546								
4 Analytical-Logical Reasoning (Type 2)	.527	.435	.457							
5 Verbal Total	.681	.519	.472	.492						
6 Quantitative Total	.614	.617	.617	.457	.472					
7 UGPA - 4 Year	.259	.209	.220	.208	.305	.195				
8 UGPA - Major Field	.272	.237	.244	.200	.290	.209	.726			
9 UGPA - Last 2 Years	.259	.224	.234	.196	.276	.181	.684	.691		
10 EXPERIMENTAL Supporting Conclusions	.628	.516	.516	.459	.596	.544	.237	.228	.221	
Number of Items	40	15	11	4	80	55				20
Raw Score Mean	19.94	10.14	5.79	1.83	34.14	25.46				7.68
Raw Score S.D.	9.36	4.04	3.02	1.24	13.86	11.14				4.25

GRE BOARD RESEARCH REPORTS OF A TECHNICAL NATURE

Boldt, R. R. Comparison of a Bayesian and a Least Squares Method of Educational Prediction. 70-3P, 1975.

Campbell, J. T. and Belcher, L. H. Word Associations of Students at Predominantly White and Predominantly Black Colleges. 71-6P, 1975.

Campbell, J. T. and Donlon, T. F. Relationship of the Figure Location Test to Choice of Graduate Major. 75-7P, 1980.

Carlson, A. B.; Reilly, R. R.; Mahoney, M. H.; and Casserly, P. L. The Development and Pilot Testing of Criterion Rating Scales. 73-1P, 1976.

Carlson, A. B.; Evans, F.R.; and Kuykendall, N. M. The Feasibility of Common Criterion Validity Studies of the GRE. 71-1P, 1974.

Centra, J. A. Graduate Degree Aspirations of Ethnic Student Groups Among GRE Test-Takers. 77-7P, 1980.

Donlon, T. F. An Exploratory Study of the Implications of Test Speededness. 76-9P, 1980.

Donlon, T. F.; Reilly, R. R.; and McKee, J. D. Development of a Test of Global vs. Articulated Thinking: The Figure Location Test. 74-9P, 1978.

Echternacht, G. Alternate Methods of Equating GRE Advanced Tests. 69-2P, 1974.

Echternacht, G. A Comparison of Various Item Option Weighting Schemes/A Note on the Variances of Empirically Derived Option Scoring Weights. 71-17P, 1975.

Echternacht, G. A Quick Method for Determining Test Bias. 70-8P, 1974.

Evans, F. R. The GRE-Q Coaching/Instruction Study. 71-5aP, 1977.

Fredericksen, N. and Ward, W. C. Development of Measures for the Study of Creativity. 72-2P, 1975.

Kingston, N. and Dorans, N. Effect of the Position of an Item Within a Test on Item Responding Behavior: An Analysis Based on Item Response Theory. 79-12bP, 1982.

Kingston, N. M. and Dorans, N. J. The Feasibility of Using Item Response Theory as a Psychometric Model for the GRE Aptitude Test. 79-12P, 1982.

Levine, M. V. and Drasgow, F. Appropriateness Measurement with Aptitude Test Data and Estimated Parameters. 75-3P, 1980.

McPeek, M.; Altman, R. A.; Wallmark, M.; and Wingersky, B. C. An Investigation of the Feasibility of Obtaining Additional Subscores on the GRE Advanced Psychology Test. 74-4P, 1976.

Pike, L. Implicit Guessing Strategies of GRE Aptitude Examinees Classified by Ethnic Group and Sex. 75-10P, 1980.

Powers, D. E.; Swinton, S.; Thayer, D.; and Yates, A. A Factor Analytic Investigation of Seven Experimental Analytical Item Types. 77-1P, 1978.

Powers, D. E.; Swinton, S. S.; and Carlson, A. B. A Factor Analytic Study of the GRE Aptitude Test. 75-11P, 1977.

Reilly, R. R. and Jackson, R. Effects of Empirical Option Weighting on Reliability and Validity of the GRE. 71-9P, 1974.

Reilly, R. R. Factors in Graduate Student Performance. 71-2P, 1974.

Rock, D. A. The Identification of Population Moderators and Their Effect on the Prediction of Doctorate Attainment. 69-6bP, 1975.

Rock, D. A. The "Test Chooser": A Different Approach to a Prediction Weighting Scheme. 70-2P, 1974.

Rock, D., Werts, C., and Grandy, J. Construct Validity of the GRE Aptitude Test Across Populations--An Empirical Confirmatory Study. 78-1P, 1982.

Sharon, A. T. Test of English as a Foreign Language as a Moderator of Graduate Record Examinations Scores in the Prediction of Foreign Students' Grades in Graduate School. 70-1P, 1974.

Stricker, L. J. A New Index of Differential Subgroup Performance: Application to the GRE Aptitude Test. 78-7P, 1981.

Swinton, S. S. and Powers, D. E. A Factor Analytic Study of the Restructured GRE Aptitude Test. 77-6P, 1980.

Ward, W. C. A Comparison of Free-Response and Multiple-Choice Forms of Verbal Aptitude Tests. 79-8P, 1982.

Ward, W. C.; Frederiksen, N.; and Carlson, S. B. Construct Validity of Free-Response and Machine-Scorable Versions of a Test of Scientific Thinking. 74-8P, 1978.

Ward, W. C. and Frederiksen, N. A Study of the Predictive Validity of the Tests of Scientific Thinking. 74-6P, 1977.